Search for Pair Production of Second Generation Scalar Leptoquarks in $p\bar{p}$ Collisions at the Tevatron

Dissertation der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von
Philippe Calfayan
geboren in Argenteuil, Frankreich

München, den 7. Mai 2008

Search for Pair Production of Second Generation Scalar Leptoquarks in $p\bar{p}$ Collisions at the Tevatron

Dissertation der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von
Philippe Calfayan
geboren in Argenteuil, Frankreich

München, den 7. Mai 2008

1. Gutachterin / 1^{st} referee : Prof. Dr. Dorothee Schaile
2. Gutachter / 2^{nd} referee : Prof. Dr. Martin Faessler

Tag der mündlichen Prüfung / date of oral exam : 9. Juli 2008

Abstract

As predicted by numerous extensions of the Standard Model, leptoquarks (LQ) are hypothetical bosons allowing lepton-quark transitions. Under the assumption that they couple only to quarks and leptons of the same generation, three generations of leptoquarks can be distinguished.

The search for the pair production of second generation scalar leptoquarks has been carried out in $p\overline{p}$ collisions at $\sqrt{s} = 1.96\,\text{TeV}$, using an integrated luminosity of $1\,\text{fb}^{-1}$ collected by the DØ experiment at the Tevatron collider between August 2002 and February 2006. Topologies arising from the $LQ\overline{LQ} \to \mu q \nu q$ and $LQ\overline{LQ} \to \mu q \mu q$ decay modes have been investigated. In order to maximize the available statistics, a method for the combination of various prescaled triggers with an inclusive OR has been developed.

Since no excess of data over the Standard Model prediction has been observed, upper limits on the leptoquark pair production cross section have been derived at 95% confidence level as function of the leptoquark mass and the branching fraction $\beta = Br(LQ \to \mu q)$, and are interpreted as lower limits on the leptoquark mass as function of β.

For $\beta = 1$, $\beta = 1/2$ and $\beta = 0.1$, the combination of the two channels excludes scalar second generation leptoquarks with masses up to 309 GeV, 262 GeV, and 174 GeV, respectively. The lower bounds on the scalar second generation leptoquark mass obtained for $\beta \geq 0.1$ are the best exclusion limits to date.

Zusammenfassung

Viele Erweiterungen des Standardmodells sagen die Existenz von zusätzlichen Bosonen, sogenannten Leptoquarks (LQ), hervor, die Übergänge zwischen Quarks und Leptonen ermöglichen. Diese Arbeit beschreibt die Suche nach der Paarproduktion von skalaren Leptoquarks der zweiten Generation in Proton-Antiproton-Streuung bei einer Schwerpunktsenergie von $\sqrt{s} = 1.96\,\text{TeV}$ in den Produktions- und Zerfallskanälen $LQ\overline{LQ} \to \mu q \nu q$ und $LQ\overline{LQ} \to \mu q \mu q$. Hierbei wurden die entsprechenden Ereignistopologien in einem Datensatz mit einer integrierten Luminosität von $1\,\text{fb}^{-1}$, der mit dem DØ-Detektor am Tevatron Speicherring aufgezeichnet wurde, untersucht. Um die zur Verfügung stehende Ereignisstatistik zu erweitern, wurde eine Methode entwickelt, die verschiedene Myontrigger kombiniert. Es wurde kein Überschuss an Daten im Vergleich zur Vorhersage des Standardmodells beobachtet. Somit wurden obere Grenzen für den Wirkungsquerschnitt der Leptoquark-Paarproduktion als Funktion der Masse des hypothetischen Leptoquarks M_{LQ} und des angenommenen Verzweigungsverhältnis $\beta = Br(LQ \to \mu q)$ bestimmt und damit untere Schranken auf M_{LQ} als Funktion von β ermittelt. Die Kombination beider Kanäle liefert untere Ausschlussgrenzen für M_{LQ} von 309 GeV, 262 GeV und 174 GeV bei $\beta = 1$, 0.5, beziehungsweise 0.1. Die ermittelten Ausschlussgrenzen für skalare Leptoquarks der zweiten Generation sind für alle $\beta \geq 0.1$ signifikant im Vergleich zu vorhergehenden Experimenten erweitert worden.

Acknowledgments

I would first like to thank my parents and brothers for their support, patience and love as I have been staying abroad, in Munich and at Fermilab.

Regarding the thesis itself, I would like to thank Michael Klasen for his devotion to helping me in the search for a PhD position abroad, Dorothee Schaile for accepting me in the Munich group and for her confidence, Otmar Biebel for his support, and Thomas Nunnemann for his availability and help, especially during the critical stages of the work. The achievement of this thesis would not have been possible without the additional help of other members of the Munich group, particularly Raimund Ströhmer. The contribution of the whole DØ collaboration is also to be stressed, and I do not forget the assistance of Boris Tuchming and Sabine Lammers.

My integration into the German landscape has been significantly eased by the friendship of the Munich group members, especially Britta and Alexander. Thanks to the conviviality of the other German groups, I could also feel at home during my stay at Fermilab.

Last but not least, I would also like to thank my friends of the early days for their support and their faith.

Contents

Introduction

The field of Elementary Particle Physics is, to date, best described by what is referred to as the Standard Model (SM) [1], in which gravitation is omitted. Although the Standard Model predictive power proves to be of a tremendous accuracy at the scale of electroweak interactions ($\mathcal{O}(100$ GeV)), theoretical insufficiencies appearing at higher energies ($\geq 1\,$TeV) imply that it cannot be considered as the universal theory of all interactions, even if one neglects gravity.

Among the inelegancies of the Standard Model, the observed symmetry between leptons and quarks, which both come in three generations, is not explained. Extensions of the Standard Model, which try to unify all fundamental interactions, naturally tend to introduce new particles with masses above ~ 100 GeV [1]. The symmetry between leptons and quarks can be explained by considering new gauge bosons with couplings between the lepton and quark sector. Such particles, commonly referred to as leptoquarks (LQ) [2–6], would carry both lepton and baryon quantum numbers, as well as a fractional electric charge. The simplest extensions of the Standard Model, based on a larger gauge symmetry group, usually predict the existence of highly massive leptoquarks which enter in contradiction with experimental measurements.

In 1997, the H1 and ZEUS collaborations, working at the *ep* collider HERA, simultaneously observed an excess of neutral and charged current events at high four-momentum transfer [7, 8]. Incompatible with the Standard Model, these events of electron+jet topology were presenting a reconstructed electron-jet mass of about 200 GeV, which could be explained by the single production of a leptoquark followed by its decay into an electron and a quark which hadronizes into a jet. Such a low-mass leptoquark would be produced at high rates by hadron colliders like Tevatron, but was ruled out by the experiments CDF and DØ experiment [9].

In order to provide an underlying theory to the search of such low-energy leptoquarks, investigations are carried out within the scope of an effective model, which is by construction consistent with the Standard Model. Experimental bounds on the proton decay and on flavor changing neutral currents constrain the hypothetical leptoquarks to couple only to quarks and leptons of the same generation, which leads to three generations of leptoquarks, distinguished by the fermion generation of their decay products.

Searches for leptoquarks at the Tevatron have been further carried on. Until the Large Hadron Collider (LHC) starts running, Tevatron will remain the best tool for such investigations. Using $1\,\text{fb}^{-1}$ of integrated luminosity collected by the DØ experiment at the Tevatron between August 2002 and February 2006, a search for the pair production of second generation scalar leptoquarks is described in this document in the case of the two channels arising from the $LQ\overline{LQ} \rightarrow \mu q \nu q$ and $LQ\overline{LQ} \rightarrow \mu q \mu q$ decay modes.

In the first chapter, the limitations of the Standard Model are briefly highlighted. Extended models, which cope with part of these insufficiencies, are presented, with focus on the way how

[1] This document assumes $c = \hbar = 1$.

leptoquark are naturally introduced. The effective model on which the presented search relies is described, and the resulting phenomenology at the Tevatron is discussed. The case of the two channels analyzed in this document is emphasized.

The second chapter consists in the description of the experimental apparatus with which the search for leptoquarks is carried out. The Tevatron hadron collider and the DØ detector are thus presented. The reconstruction of the particles, as seen by the detector, and the quality requirements imposed to them are detailed in the third chapter, which begins with the description of the samples considered in the analysis.

The efficiency of data taking plays a crucial role in new phenomena searches. The various triggers, which achieve the selection of specific physics events, can be combined in order to increase the statistics available for the analysis. Due to bandwidth limitations, trigger rates can be limited artificially, which entails significant complications for such a combination. The fourth chapter focuses on a method developed in order to manage the combination of multiple single muon triggers, which enables the selection of final states containing muons as required for the analysis presented.

The leptoquark decay modes investigated in this thesis are leading to final states similar to those of Standard Model processes, which are therefore treated as a background. Standard Model predictions are simulated in order to be compared to the collected data. Leptoquark pair production candidate events are first selected with a simple analysis based on kinematic requirements and reconstructed variables. A more complicated but more efficient selection is obtained by considering a multivariate classifier. All the steps of the candidate events selection are provided in the fifth chapter.

Since no excess of data over the Standard Model prediction is observed, upper limits on the leptoquark pair production cross section are derived at 95% confidence level as function of the leptoquark mass and the branching fraction $\beta = Br(LQ \to \mu q)$ for the leptoquark to decay to a muon and a quark. The last chapter describes the calculation of the cross section upper limits, and their interpretation as lower bounds on the leptoquark mass depending on β. The results are compared with the preceding searches for leptoquarks.

Chapter 1

Theoretical motivation

This chapter briefly introduces the Standard Model of Particle Physics, and stresses its limitations. Numerous extensions of the Standard Model include a symmetry between the quark and lepton sectors and thus naturally include leptoquarks as new particles. Under the assumptions of an effective model, leptoquarks can be investigated at hadron colliders. The possible production modes and signatures are discussed, and the channels analyzed in this thesis are detailed.

1.1 The Standard Model

1.1.1 Particles and interactions

The Standard Model of strong [10–12], electromagnetic and weak forces is a chiral gauge theory relying on Quantum Field Theory. Electromagnetic and weak interactions are unified under the $SU(2)_L \times U(1)_Y$ electroweak [13–15] symmetry[1], and Quantum Chromodynamics (QCD), which describes strong interactions, is based on the $SU(3)_C$ symmetry group[2]. The Standard Model does not attempt to describe gravitation.

Interactions	electromagnetic	weak	strong
Gauge group	$U(1)_Y$	$SU(2)_L$	$SU(3)_C$
Mediator	photon γ	$W^{\pm}$, Z	gluon g
Coupling constant	1/137	10^{-6}	< 1
Range (m)	∞	10^{-18}	$\leq 10^{-15}$

Table 1.1: *Fundamental interactions, and their associated gauge group, mediators (spin-1 bosons), and properties.*

The Standard Model particles are described as field excitations carrying charges which remain conserved under the predicted interactions. Interactions, listed in Table 1.1, are mediated through spin-1 bosons, which arise from the invariance of the Standard Model Lagrangian under $SU(3)_C \times SU(2)_L \times U(1)_Y$ local gauge transformations. The matter is build of spin-$\frac{1}{2}$ fermions, which come in three families. The absence of right-handed neutrino implies the chirality of the electroweak interactions, which acts on left-handed weak isospin doublets, as represented in Table 1.2.

[1] The subscripts L and Y refer to the left chirality of the interactions and the electroweak hypercharge, respectively.

[2] The subscript C refers to the quark color.

	Fermion generations		
leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$ e_R	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$ μ_R	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$ τ_R
quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$ $u_R\,,\, d_R$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$ $c_R\,,\, s_R$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$ $t_R\,,\, b_R$

Table 1.2: *Quarks and leptons (spin-$\frac{1}{2}$ fermions) of the Standard Model. The subscripts R and L denotes the left and right chiralities, respectively. Left particles are grouped in weak isospin doublet. The anti-particles, not shown, have the same properties except an opposite electromagnetic charge, and an opposite chirality.*

Because of symmetry constraints of the Standard Model gauge groups, fermions and bosons are predicted to be massless. The masses of the $W^\pm$ and Z bosons are assumed to be generated by the Higgs mechanism [16], which consists in a spontaneous breaking of the electroweak $SU(2)_L \times U(1)_Y$ symmetry down to the $U(1)$ symmetry. The Higgs mechanism introduces an hypothetical scalar field possessing a non-zero vacuum expectation value due to its self-interaction. The Yukawa couplings between the fermions and the Higgs field give masses to the fermions after electroweak symmetry breaking. A consequence of the Higgs mechanism is the existence of the scalar Higgs boson, which has not been discovered so far.

1.1.2 Success and limitations

Many processes of the Standard Model can be calculated and confronted to experiments. Unpredicted parameters (like the fermion masses) can thus be measured, and deviations to predictions can be investigated.

The confirmations of the Standard Model began in 1973 with the discovery of the neutral currents, and continued in 1979 and 1983, with respectively, the discoveries of the gluon and the electroweak bosons ($W^\pm$ and Z). Between 1975 and 2000, the third fermion family was experimentally confirmed. Over the last decade, quantitative and successful tests of the Standard Model have been carried out at the LEP, Tevatron, SLC and HERA particle colliders.

Although the Standard Model remarkably withstood experimental challenges, it does not appear to be the universal theory of all interactions, Even below the Planck scale $M_{Planck} = 10^{19}\,\text{GeV}$, at which the gravitation is no more negligible, the description provided by the Standard Model is not completely unified, since only the electromagnetic and weak interactions are described as stemming from a common origin. Moreover, the coupling constants associated to the electromagnetic, weak and strong interactions, extrapolated at higher energy with the renormalization group equations, are not converging perfectly.

In addition, the Standard Model turns out to be limited to energies below approximately 1 TeV. Therefore, it tends to be interpreted as an effective theory valid at the electroweak scale ($\mathcal{O}(100\text{ GeV})$). This is illustrated by the hierarchy problem, which occurs when correcting the Higgs boson mass by including higher orders radiative processes to its propagator. The quadratically divergent corrections, which arise, cannot be absorbed by the renormalization procedure,

and require to set a cut-off on the integral summing over the momenta of the particles contributing to the loop processes. A cut-off of around 1 TeV is imposed, since higher values would lead to an over-massive Higgs boson which would not be consistent with the observed phenomena at the electroweak scale.

Furthermore, the Standard Model does not provide any candidate for Dark Matter, and does not explain the baryonic asymmetry of the universe, which requires more sources of CP violation.

Contrasting with the universality expected from a completely unified theory, the Standard Model includes a large number of free parameters, which need to be evaluated by experimental measurements. Up to 26 parameters are not constrained by the theory. 7 of these parameters are introduced for the neutrinos mixing, which gives neutrinos a mass, and can thus be interpreted as a phenomenon outside the scope of the Standard Model.

Besides, if some characteristics the Standard Model remain unexplained, like the number of fermion families, others are even introduced circumstantially, such as the ad-hoc shape of the yet experimentally unproved Higgs field potential. The fact that quarks can carry up to three different colors and fractional electric charges is, as well, not predicted. Nevertheless, it turns out to be essential for the renormalizability of the theory, since it allows the exact cancellation of gauge anomalies caused by triangle diagrams, by counteracting the contribution of the leptons [17].

A symmetry between leptons and quarks is however not included in the Standard Model, since there is no direct coupling between quark and lepton families. Theories beyond the Standard Model, which try to cope with the described limitations, usually predict new interactions arising from a larger symmetry group. By gathering leptons and quarks in the same multiplets, such theories naturally allow lepton-quark transitions, which are mediated by new bosons usually called leptoquarks [2–6].

1.2 Leptoquarks and theories beyond the Standard Model

Leptoquarks are hypothetical colored bosons which carry both lepton and baryon quantum numbers and thus mediate lepton-quark transitions. Such particles are predicted in numerous extensions of the Standard Model. In the following, we briefly describe common theories beyond the Standard Model, in which new interactions between lepton and quarks naturally arise.

1.2.1 Grand Unified Theories

Part of the limitations of the Standard Model can be overcome by considering new theories based on a larger gauge symmetry group G. Such models are referred to as Grand Unified Theories (GUT) [18]. The group G is chosen such that the additional symmetries could explain the arbitrary features of the Standard Model.

In order to be consistent with experimental observations, G must include the Standard Model symmetry group (which is of rank[3] 4), to which it can be reduced after spontaneous symmetry breaking. By choosing a gauge group based on a single coupling strength, the unification of all the Standard Model interactions is made possible. The energy scale where the coupling constants of the Standard Model could converge is very large, since the observed couplings are significantly different at the electroweak scale and are varying logarithmically with energy. Thus, the Standard Model would be a low energy approximation of GUTs at the electroweak scale.

[3]Physically, the rank of a group is the total number of quantum numbers which describes a state.

The gauge group G is required to include complex irreducible representations which can accommodate the particles of the Standard Model. By gathering a fermion family within the same multiplet, a connection between lepton and quarks can be introduced, as well as mediators (bosons) for this connection. Among the groups of rank 4 involving only one coupling strength[4], SU(5) is the only one meeting the requirements of a GUT.

The simplest model based on $SU(5)$ was proposed by Georgi and Glashow [2] in 1973. This model includes $SU(3)_C \times SU(2)_L \times U(1)_Y$ as a subgroup, and can accommodate a complete fermion family with the association of the complex representations $\bar{5}$ and 10. Translated in the representation $SU(3)_C \times SU(2)_L$, $\bar{5}$ and 10 become:

$$\bar{5} = (\bar{3},1) + (1,2) = \bar{d} + \begin{pmatrix} \nu_l \\ \ell^- \end{pmatrix} \tag{1.1}$$

$$10 = (\bar{3},1) + (3,2) + (1,1) = \bar{u} + \begin{pmatrix} u \\ d \end{pmatrix} + \ell^+ \tag{1.2}$$

where u and d are the up-type and down-type quarks, respectively, and $\binom{\nu_l}{\ell^-}$ the weak isospin doublet.

The $SU(5)$ group contains 24 generators, including those of the Standard Model, namely the eight Gell-Mann matrices of $SU(3)_C$, the three Pauli matrices of $SU(2)_L$, and the phase of $U(1)_Y$. Since these generators must be represented by traceless matrices (as they are hermitian operators), and since the electric charge operator is one of them, the sum of the charges of the particles included in the representation $\bar{5}$ must vanish. This leads to the following equation:

$$3\,Q_{\bar{d}} + Q_{\ell^-} = 0 \tag{1.3}$$

where $Q_{\bar{d}}$ and Q_{ℓ^-} are the electric charges of a down-type quark and a negatively charged lepton, respectively. The preceding relation thus explains why quarks carry a fractional charge.

The 24 generators correspond to as many bosons, which are included in the 24 adjoint representation. Decomposed in the representation $SU(3)_C \times SU(2)_L$, 24 becomes:

$$24 = (3,2) + (\bar{3},2) + (8,1) + (1,3) + (1,1) \tag{1.4}$$

where $(8,1)$ includes the eight gluons, $(1,3) + (1,1)$ gathers the electroweak bosons (γ,Z,$W^{\pm}$). $(3,2)$ and $(\bar{3},2)$ contains twelve new bosons, 6 X and 6 Y, of electric charge $\pm\frac{4}{3}$ and $\pm\frac{1}{3}$ respectively, and which are sensible to both $SU(3)_C$ and $SU(2)_L$ gauge transformations.

The new X and Y bosons can mediate lepton-quark or quark-quark transitions, which violate the conservation of the lepton (L) and baryon (B) numbers while conserving $B - L$. The most popular example of the consequences of such quark-quark transitions is decay of the proton into mesons and leptons, as represented in Figure 1.1. Since its experimental lower limit is $\tau_p > 10^{31}$ years [19], high lower bound on the mass of the X and Y bosons are set, namely $M \gtrsim 10^{15}$ GeV. Such a scale is yet not accessible by experiments. Propagated down to the electroweak scale, the predictions of the Georgi and Glashow $SU(5)$ model are not in agreement with experimental data.

[4] These groups are $Sp(8)$, $SO(8)$, $SO(9)$, F_4, $SU(3) \times SU(3)$ and $SU(5)$. The four first cited have only real representations, and $SU(3) \times SU(3)$ cannot accommodate both integrally and fractionally charged particles.

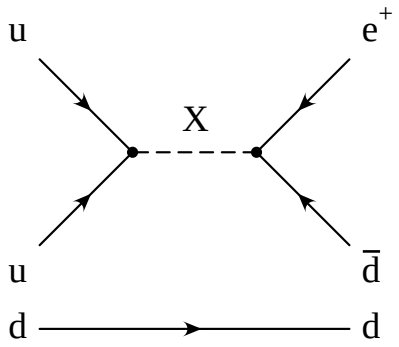

Figure 1.1: *Decay of the proton into a positron and neutral pion (π^0), through the mediation of the X boson. This process violates both the lepton and baryon numbers.*

Other GUT models including one coupling constant and based on gauge groups of rank greater than 4 have been postulated. For instance, the $SO(10)$ group, of rank 5, allows to build a theory where all the fermions are included within the same irreducible representation, namely $\overline{16}$. Several ways to break $SO(10)$ down to the electroweak scale are possible, one of them leading to the Pati-Salam GUT model [20], of group $SU(4)_c \times SU(2)_L \times SU(2)_R$, and which interprets the lepton sector as the fourth color. Models based on even larger symmetry group, like on $SU(15)$, which can also gather all the fermions within its fundamental representation 15, have however the drawback of introducing numerous new parameters.

1.2.2 Supersymmetry

Supersymmetry (*SUSY*) [21] allows the unification of matter and interactions. A supersymmetric transformation acts on the spin of a particle, and associates a fermion to a boson, and a boson to a fermion. The new particles arising from this symmetry are referred to as superpartners. It is worth mentioning that the introduction of superpartners cancels the divergences of the Higgs mass corrections, as their contribution counteract that of the Standard Model particles. This is due to the fact that loop corrections have amplitudes of opposite signs depending on whether the contributing particle is a boson or a fermion. Since no supersymmetric particle has yet been observed, the supersymmetry must be broken. Various symmetry breaking scenarios are possible, leading to numerous supersymmetric models. Furthermore, the precise convergence of the three coupling constants contributes to make *SUSY* models serious candidates for an universal theory of all interactions.

The simplest supersymmetric extension of the Standard Model, which introduces a minimal addition of superpartners, is referred to as the Minimal Supersymmetric Standard Model (MSSM) [22]. In the MSSM, the Higgs potential is constrained, and thus no more arbitrary.

In order to explain the stability of the proton in the MSSM, a new quantum number is introduced, the *R-parity* (R), as defined by the following relation:

$$R = (-1)^{3B+2S+L} \tag{1.5}$$

where S is the spin of the particle, and B and L the baryon and lepton numbers, respectively.

The *R-parity* is discrete and multiplicative. It is equal to 1 for Standard Model particles, and -1 for supersymmetric particles. Although *R-parity* is supposed to be conserved, its violation is

theoretically allowed, but yet leads to contrasting phenomenologies.

For a given process, the conservation of *R-parity* is achieved in the same way as in the case of spatial parity. The initial and final parities are equal to the product of the parities of the initial and final states, respectively. As a result, when *R-parity* is conserved, *SUSY* particles can only be produced in pairs, and can decay only to another *SUSY* particle in association with a particle from the Standard Model. The lightest *SUSY* particle (LSP), arising from successive *SUSY* particle decays, must thus be stable. A Neutral LSP thus constitute a serious candidate for dark matter [23].

In the case of violated *R-parity* [24], *SUSY* particles are allowed to decay exclusively into Standard Model particles, and by considering the opposite process, the single production of *SUSY* particles is thus also allowed. As a consequence, the lightest *SUSY* particle is no more stable, and cannot describe dark matter.

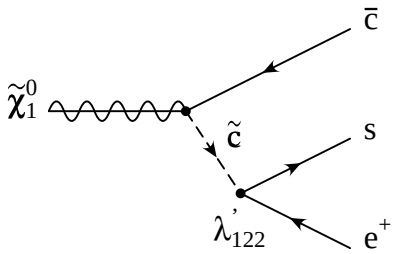

Figure 1.2: *Decay of the neutralino (the supersymmetric partner of Standard Model neutral bosons) into an electron and two quarks. R-parity is assumed to be violated. In the process, the supersymmetric partner of the quark (the squark) is decaying to an electron and a quark through a leptoquark-like coupling.*

Assuming the violation of *R-parity*, supersymmetric bosons are thus allowed to decay to Standard Model fermions, by violating the lepton and baryon numbers, and involving a leptoquark-like coupling. Figure 1.2 represents an example of a *R-parity* violated process, namely the decay of a neutralino (the superpartner of Standard Model neutral bosons) to a squark and a quark (which conserves *R-parity*), followed by the decay of the squark to an electron and a quark (the superpartner of a quark), which violates *R-parity* through a leptoquark-like coupling (λ'_{122}). Experimental and theoretical constraints on the leptoquark-lepton-quark coupling can thus be translated into constraints on the squarks [25].

1.3 Leptoquark effective models

The search for leptoquarks can be carried out by relying on effective models, and thus independently of theories beyond the Standard Model such as those presented in Section 1.2. However, constraints on the nature of the leptoquarks and on their interactions with Standard Model particles have to be imposed so as to be consistent with experimental observations. The following description is based on Reference [25].

A general effective leptoquark model, as proposed by Buchmüller, Rückl and Wyler (BRW) [26], requires leptoquarks to have normalizable interactions, to obey Standard Model gauge groups ($SU(3)_C \times SU(2)_L \times U(1)_Y$) symmetries, and to couple only to Standard Model fermions and gauge bosons.

In order to ensure the proton stability, leptoquarks are further constrained to conserve the lepton and baryon numbers separately. This leads to seven scalar and seven vector leptoquarks, carrying the fermionic number $F = 3\,B + L$ of either $|F| = 0$ or $|F| = 2$.

Table 1.3 lists the resulting seven scalar (S) and seven vector (V) leptoquarks, along with their electric and weak charges, their fermionic numbers, and their decay products. Within each isospin family, mass degeneracy is assumed for experimental searches, which is theoretically motivated when omitting loop corrections. The symbols referring to the leptoquarks in Table 1.3 therefore designates any of its electric charge state. In addition, the L and R subscripts on the leptoquark symbol denotes to the chirality of the coupled lepton. For instance, the symbol $S_{\frac{1}{2},L}$ both refers to a scalar leptoquark of electric charge $-\frac{5}{3}$ and $-\frac{2}{3}$, which couples to a left-handed lepton. weak isospin is appended as a subscript, and is equal to 0, $\frac{1}{2}$, or 1.

Although all 14 leptoquarks appear in a GUT based on $SU(15)$, only a subset of the BRW lepoquarks are generally included in a specific fundamental model (for example, V_0 is part of the Pati-Salam GUT model).

The branching fractions of the leptoquark decays into a charged lepton and quark or neutrino and quark ($\beta = Br(LQ \to lq)$) are determined by the respective leptoquark-lepton-quark coupling. By construction, it can thus only take the discrete values 0, $\frac{1}{2}$, or 1.

Under the preceding assumptions, leptoquarks could, in principle, decay into any combination of a quark and a lepton, but leptoquarks with masses as low as $\mathcal{O}(100\text{ GeV})$ are only allowed to couple to one generation of quarks and leptons, since they otherwise would generate lepton number violation and sizable flavor-changing neutral currents. An example of flavor-changing neutral current mediated by a leptoquark is provided in Figure 1.3, where the decay of the B^0 meson into a tau and a anti-muon is represented. Under this effective model, leptoquarks can thus be indexed by the generation of the fermion to which it couples. There are therefore three generations of leptoquarks.

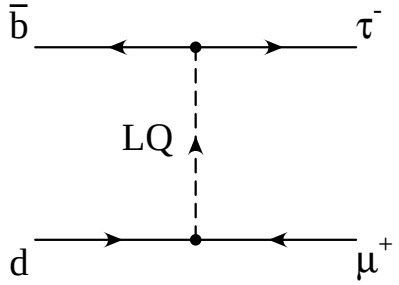

Figure 1.3: *Example of flavor-changing neutral current. A B^0 meson is decaying into a tau and an anti-muon.*

Additional low energy constraints, such as chirally suppressed meson decays (e.g. $\pi \to e\nu$) or virtual-loop contributions to the $g-2$ of the muon, entails the necessity of assuming that leptoquarks have pure chiral couplings to Standard Model fermions. Augmented by the two previous requirements, the BRW model is called the "minimal Buchmüller-Rückl-Wyler effectiv model" (mBRW) [26].

The interaction of the leptoquarks, as defined under the mBRW model, with the Standard Model bosons is completely determined in the case of scalar leptoquarks, since it only relies on Standard Model interactions. The coupling between vector leptoquarks and Standard Model

LQ type	$\|F\|$	Q	T_3	Decay products
$S_{0,L}$	2	$-1/3$	0	$\ell_L^- u_L, \nu_L d_L$
$S_{0,R}$	2	$-1/3$	0	$\ell_R^- d_R$
$\tilde{S}_{0,R}$	2	$-4/3$	0	$\ell_R^- d_R$
$S_{1,L}$	2	$-4/3$	-1	$\ell_L^- d_L$
	2	$-1/3$	0	$\ell_L^- u_L, \nu_L d_L$
	2	$+2/3$	$+1$	$\nu_L u_L$
$S_{\frac{1}{2},L}$	0	$-5/3$	-1/2	$\ell_L^- \bar{u}_L$
$S_{\frac{1}{2},R}$	0	$-5/3$		$\ell_R^- \bar{u}_R$
	0	$-2/3$	$+1/2$	$\ell_R^- \bar{d}_R$
$\tilde{S}_{\frac{1}{2},L}$	0	$-2/3$	-1/2	$\ell_L^- \bar{d}_L$
	0	$+1/3$	$+1/2$	$\nu_L \bar{d}_L$
$V_{\frac{1}{2},L}$	2	$-4/3$	-1/2	$\ell_L^- d_R$
$V_{\frac{1}{2},R}$	2	$-4/3$		$\ell_R^- d_L$
	2	$-1/3$	$+1/2$	$\ell_R^- u_L$
$\tilde{V}_{\frac{1}{2},L}$	2	$-1/3$	-1/2	$\ell_L^- u_R$
	2	$+2/3$	$+1/2$	$\nu_L u_R$
$V_{0,L}$	0	$-2/3$	0	$\ell_L^- \bar{d}_R, \nu_L \bar{u}_R$
$V_{0,R}$	0	$-2/3$	0	$\ell_R^- \bar{d}_L$
$\tilde{V}_{0,R}$	0	$-5/3$	0	$\ell_R^- \bar{u}_L$
$V_{1,L}$	0	$-5/3$	-1	$\ell_L^- \bar{u}_R$
	0	$-2/3$	0	$\ell_L^- \bar{d}_R, \nu_L \bar{u}_R$
	0	$+1/3$	$+1$	$\nu_L \bar{d}_R$

Table 1.3: *Classification of the leptoquarks in the Buchmüller-Rückl-Wyler effective model. The content of seven scalar (S) and seven vector (V) leptoquarks is listed along with their fermionic numbers ($F = 3B + L$), their electric and weak charges (Q and the third component T_3 of the weak isospin), and their possible decay products. The symbols referring to the leptoquarks denotes any of their electric charge state, the appended L and R subscripts designate to the chirality of the coupled lepton, and the appended number is the weak isospin quantum number.*

bosons is however more complicated, as it depends on trilinear and quartic couplings, which might require the introduction of anomalous couplings, which are free parameters. Four anomalous independent coupling are associated to the electroweak sector, and two additional are introduced for the strong sector. More details on the interaction of vector leptoquarks with the Standard Model bosons are provided in References [27, 28].

The assumptions of the mBRW effective model allows relatively small leptoquark masses in reach of hadron colliders like Tevatron. Enriched phenomenology can be predicted when reasonably relaxing some of its constraints. For instance, assuming that leptoquarks not only couple to Standard Model bosons but also to other unknown new fields enables the search for generic models, in which the the branching fraction $\beta = Br(LQ \to lq)$ can take arbitrary values. It is worth mentioning that under this assumption, the leptoquarks $\tilde{S}_{\frac{1}{2}}$ and S_0 could be associated to the $\tilde{u}$ and $\tilde{d}$ squarks (superpartners of the quarks), respectively.

The search for second generation scalar leptoquarks performed in this thesis relies on the mBRW with the exception that β is assumed to be a free parameter, which allows a more general investigation.

1.4 Leptoquark phenomenology at hadron colliders

Both single and pair production of leptoquarks can occur at a hadron collider like Tevatron, i.e., in $p\bar{p}$ collisions.

The single leptoquark production depends on the unknown leptoquark-lepton-quark coupling (λ_{l-q}) and therefore relies on the theoretical model considered. Figure 1.4 shows an example of single leptoquark production, in the t-channel.

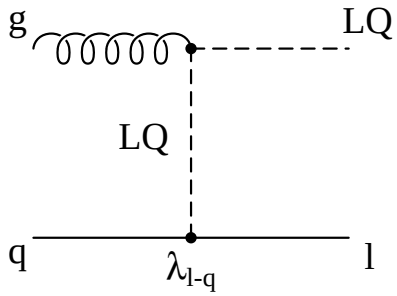

Figure 1.4: *Leptoquark single-production process in $p\bar{p}$ collisions (t-channel).*

By contrast, the leptoquark pair production is predominantly realized in the s-channel, via the strong coupling, independently of (λ_{l-q}), and thus only depends on the assumed leptoquark mass. The additional contribution to the leptoquark pair production due to t-channel lepton exchange with a cross section proportional to λ^2_{l-q} could actually also be considered, but turns out to be negligible in the case of the search for second or third generation leptoquarks. When adopting the general assumption that the leptoquark couples to the lepton and the quark of the same generation, this process is further suppressed by vanishing quark parton distribution functions (PDF) at high $Bjorken\ x$, and is thus negligible in comparison with uncertainties on the predicted cross section.

Searches for leptoquarks at electron-proton colliders (such as HERA), or e^+e^- colliders (such

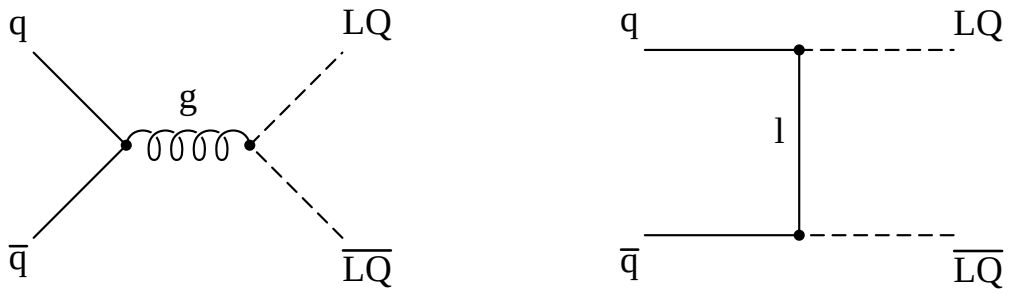

Figure 1.5: *Pair production of leptoquarks via $q\bar{q}$ annihilation at Born level. The s- (left) and t-channels (right) are represented.*

as LEP), where leptoquarks can only be singly produced, have maximized sensitivities for leptoquarks of the first generation. To the contrary, leptoquark pair production allows searches for the three generations of leptoquarks.

Reference [29] describes the calculation of the leading order (LO) and next-to-leading order (NLO) QCD cross sections for the pair production of scalar leptoquarks in $p\bar{p}$ collisions. At Born level, the leptoquark pair production can be achieved by gluon-gluon fusion, as represented in Figure 1.6, or by proton-anti-proton annihilation, as shown in Figure 1.5. Due to the dependencies in energy of the parton density functions, the leptoquark pair production via proton-anti-proton annihilation is dominant at the Tevatron ($\sqrt{s} = 1.96\,\text{TeV}$). Figure 1.7 provides the pair production cross sections for both scalar and vector leptoquarks calculated using the parton density functions (PDF) CTEQ6.1M [30], and assuming a center-of-mass energy $\sqrt{s}$ of 1.96 TeV. The vector leptoquark cross sections are derived assuming Yang-Mills or Minimal-Vector couplings, which differ by the assumed values of the anomalous couplings. Cross section for the Minimal-Vector type are smaller than in the Yang-Mills case, and both are larger than in the case of scalar leptoquarks.

At the Tevatron, leptoquark pair production can lead to three characteristic final states: $\ell^+\ell^- qq$, $\ell^\pm \nu qq$, and $\nu\nu qq$. The analysis presented in this thesis describe the search for second generation scalar leptoquark pair production in the decay modes $LQ\overline{LQ} \to \mu\nu qq$ and $LQ\overline{LQ} \to \mu\mu qq$. In the following, the branching fraction β is restrained to the case of second generation fermions, and is thus defined by $\beta = Br(LQ \to \mu q)$. The branching ratios for the three possible leptoquark pair decay modes are expressed by the formula below, as function of β:

$$Br(LQ\overline{LQ} \to \mu\mu qq) = \beta^2 \quad (1.6)$$

$$Br(LQ\overline{LQ} \to \mu\nu qq) = 2\beta(1-\beta) \quad (1.7)$$

$$Br(LQ\overline{LQ} \to \nu\nu qq) = (1-\beta)(1-\beta) \quad (1.8)$$

The branching fraction of the $LQ\overline{LQ} \to \mu\nu qq$ mode is thus maximized for $\beta = 0.5$, whereas in the $LQ\overline{LQ} \to \mu\mu qq$ mode, the best search sensitivity is obtained for $\beta = 1$. At $\beta = 0$, only the mode $LQ\overline{LQ} \to \nu\nu qq$, which is not investigated in this thesis, has a non-vanishing branching ratio.

The main background processes from the Standard Model to the pair production of leptoquarks followed by their decay into the $\mu\nu qq$ and $\mu\mu qq$ final states are the W and Z/γ^* productions, respectively. At higher energies, the production of top quark pairs also contributes significantly to these final states.

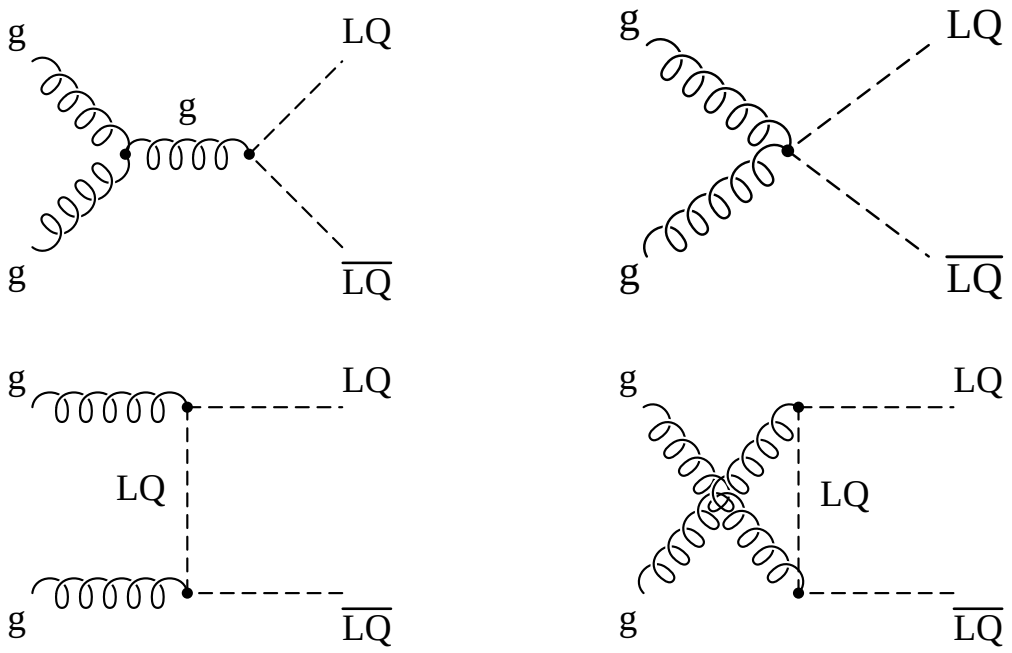

Figure 1.6: *Pair production of leptoquarks via gg fusion at Born level.*

1.5 Recent experimental results

This section summarizes the last results on leptoquark searches. More details can be found in Reference [31]. For a larger but older scope of experimental results, the reader is referred to Reference [25].

At the lepton-proton collider HERA, the limit on the leptoquark mass depends on the λ_{l-q} coupling between the leptoquark and its associated lepton and quark. As explained in the preceding section, only the search for first generation leptoquarks has a good sensitivity. Results from Hera, for leptoquarks of types $\tilde{S}_{\frac{1}{2},L}$ ($F = 0$) and $S_{0,L}$ ($F = 2$) are shown in Figure 1.8 [32], together with results from L3 and DØ (Run II).

A Tevatron, numerous direct searches for first, second and third generation leptoquarks have been carried out under the hypothesis that leptoquarks are produced in pairs. Tables 1.4 and 1.5 present the latest results (DØ and CDF) on the search for the pair production of first and second generation leptoquark. Table 1.6 provides the latest experimental lower mass limit on pair-produced third generation scalar (DØ and CDF) and vector (CDF) leptoquark.

At DØ, a search for pair-produced leptoquarks in the acoplanar jet topology ($Br(LQ \to \ell^{\pm} q) = 0$) based on $310\,\mathrm{pb}^{-1}$ (Run II) has been performed, leading to a lower mass limit of 136 GeV (calculated at 95% confidence level) [33]. This result completes the analysis presented in this thesis for $\beta = Br(LQ \to \mu^{\pm} q) = 0$, where both analyzed decay modes have no sensitivity.

The first search for single leptoquark production at hadron collider has been carried out at the DØ experiment with $300\,\mathrm{pb}^{-1}$ [34]. It assumes that the leptoquark involved in the production

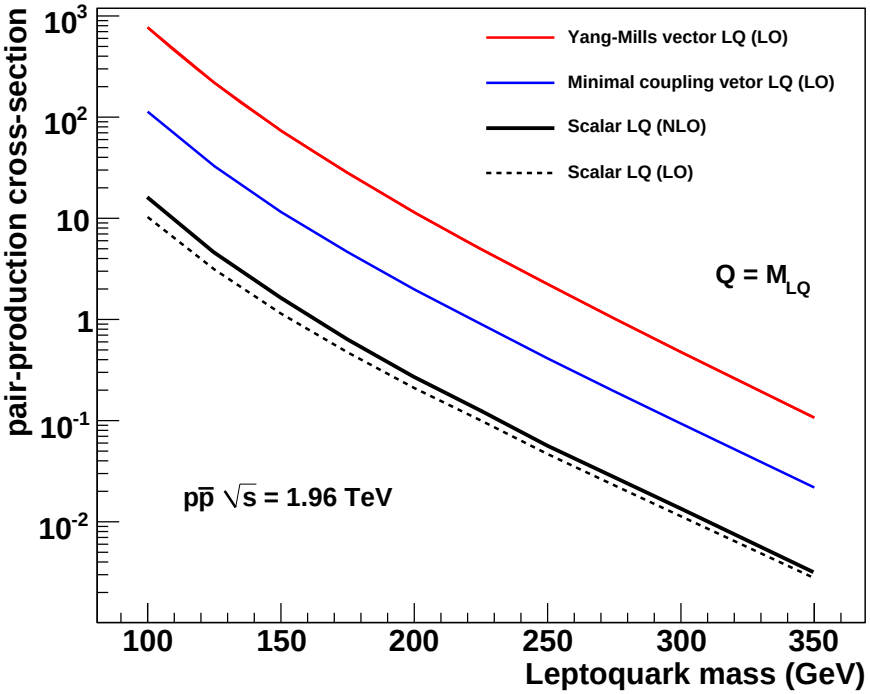

Figure 1.7: *Pair production cross section of vector leptoquarks at LO and scalar leptoquarks (in black), at LO (dashed curve) and NLO (plain curve). Two scenario are considered for the anomalous couplings, namely the Yang-Mills couplings (in red), and the Minimal-vector couplings (in blue). Calculations assume an identified factorization and renormalization scale equal to the assumed leptoquark mass. Results are provided for $p\bar{p}$ collisions at a center-of-mass energy of 1.96 TeV.*

process (see example in Figure 1.4) couples to a first generation quark, and a second generation lepton, in order to avoid suppression due to the parton density functions in the initial state. This analysis was combined with the signal bins of the leptoquark pair production search in the $LQ\overline{LQ} \to \mu\mu qq$ decay mode [35], in order to derive upper cross section limits on the production of single leptoquarks. Lower limits (calculated at 95% confidence level) on the leptoquark mass were calculated in the case $\lambda_{l-q} = 1$. Compared with the search considering only the pair production (corresponding to $\lambda_{l-q} \ll 1$), the mass limits were improved to $M_{LQ} > 274\,\text{GeV}$ assuming $\beta = Br(LQ \to \mu^{\pm} q) = 1$, and $M_{LQ} > 226\,\text{GeV}$ in the case $\beta = 0.5$.

Decay mode	Lower M_{LQ} limit (GeV)	$Br(LQ \to eq)$	Luminosity (pb^{-1})	Experiment
$eeqq$	236	1		
$+e\nu qq$	205	1/2	203 (Run II)	CDF [36]
$+\nu\nu qq$	145	0.1		
$eeqq$	256	1	250 (Run II)	DØ [37]
$+$ $e\nu qq$	234	1/2	+ 120 (Run I)	

Table 1.4: *Results on the search for the pair production of first generation scalar leptoquarks, at Tevatron. The lower limits on the leptoquark mass (M_{LQ}) are calculated at 95% confidence level.*

Decay mode	Lower M_{LQ} limit (GeV)	$Br(LQ \to \mu q)$	Luminosity (pb^{-1})	Experiment
$\mu\nu qq$	170	1/2	198 (Run II)	CDF [38]
$\mu\mu qq$	226	1		
$+$	208	1/2	198 (Run II)	CDF [38]
$\mu\nu qq$	143	0.1		
$\mu\mu qq$ (Run II)	251	1	294 (Run II)	DØ [35]
$+$ $\mu\nu qq$ (Run I)	204	1/2	94 (Run I)	

Table 1.5: *Results on the search for the pair production of second generation scalar leptoquarks, at Tevatron. The lower limits on the leptoquark mass (M_{LQ}) are calculated at 95% confidence level.*

Decay mode	Lower M_{LQ} limit (GeV)	Branching ratio	Luminosity (pb^{-1})	Experiment
$\tau\tau bb$	180	$Br(LQ \to \tau b) = 1$	1000 (Run II)	DØ [39]
$\nu_\tau\nu_\tau bb$	229	$Br(LQ \to b\nu_\tau) = 1$	425 (Run II)	DØ [40]
$\tau\tau qq$	251 (MV), 317 (YM)	$Br(LQ \to \tau b) = 1$	322 (Run II)	CDF [41]

Table 1.6: *Results on the search for the pair production of third generation scalar and vector leptoquarks, at Tevatron (Run II). The lower limits on the leptoquark mass (M_{LQ}) are calculated at 95% confidence level. For the vector leptoquarks analysis, both the Yang-Mills (YM) and Minimal-Vector (MV) couplings scenario are considered.*

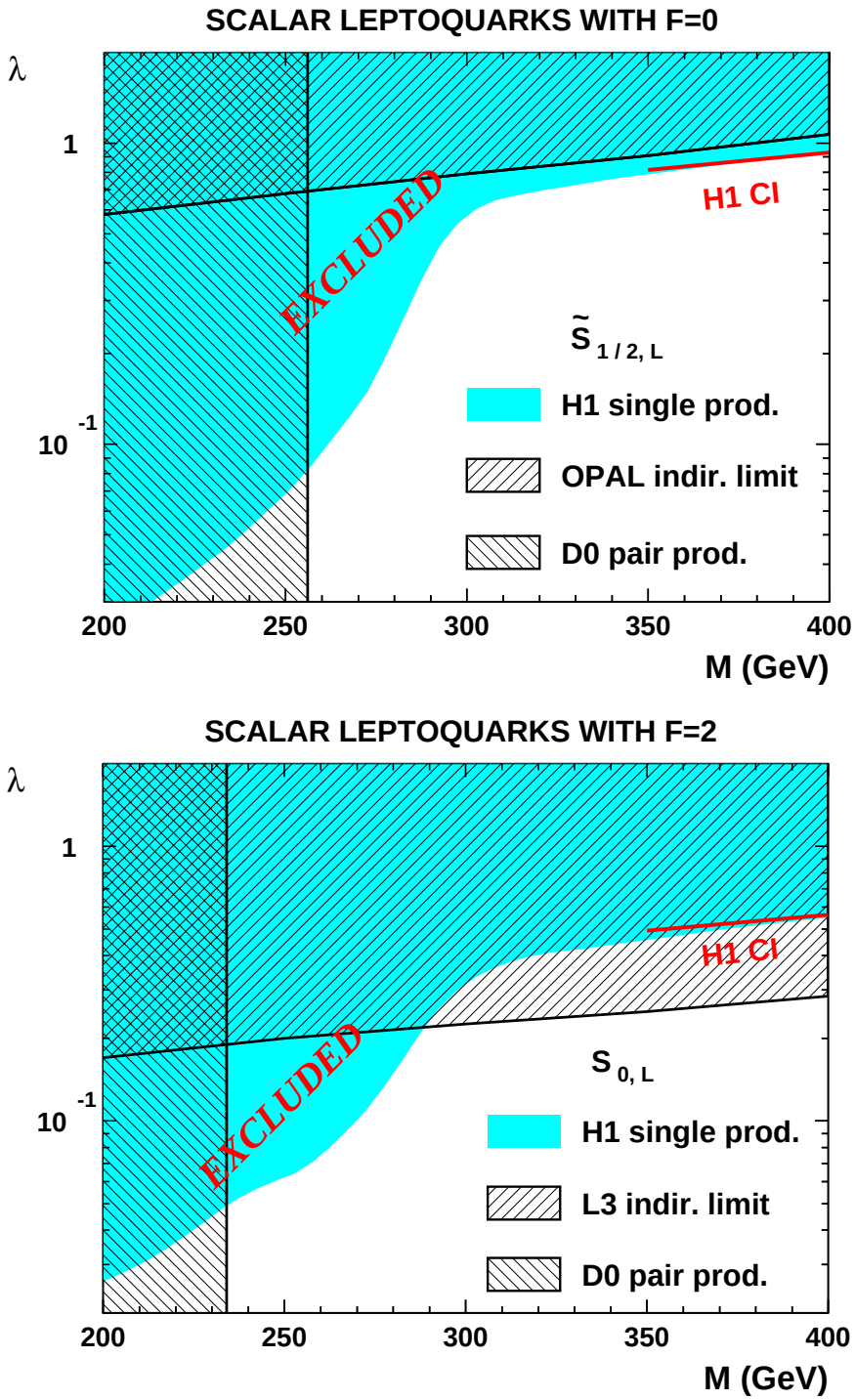

Figure 1.8: *Search for first generation leptoquarks at Hera. Mass lower limits on both scalar leptoquarks of types* $\tilde{S}_{\frac{1}{2},L}$ *(upper plot) and* $S_{0,L}$ *(lower plot) are provided as function of* λ_{l-q}*, along with results from L3 and DØ (Run II) [32]. The mass limits on* $\tilde{S}_{\frac{1}{2},L}$ *and* $S_{0,L}$ *assume* $Br(LQ\overline{LQ} \to eq)$ *equal to 1 and 0.5, respectively.*

Chapter 2

The Tevatron and the DØ detector

This chapter describes the experimental apparatus with which the presented search is carried out. The Tevatron proton-anti-proton collider, located at the Fermi National Accelerator Laboratory (FNAL), is introduced, and the DØ detector, which is used to record high energetic collisions, is presented.

2.1 The Fermilab accelerator complex

The Fermilab accelerator complex [42, 43] is composed of six synchrotron rings and two linear accelerators, where beams of protons and anti-protons are produced and accelerated. The most energetic accelerator is the Tevatron collider, which has a circumference of about 6 km, and operates collisions between protons and anti-protons moving in opposite directions. The crossing of the hadron beams occurs in two interaction regions, surrounded respectively by the CDF and the DØ detectors. A sketch of the accelerators chain is provided in Figure 2.1. Until the Large Hadron Collider, at CERN, starts running, the Tevatron will remain the particle collider with the highest center-of-mass energy.

The Tevatron activity can be separated into two main periods: Run I and Run II. The first phase took place between 1992 and 1996 and provided a center-of-mass energy of 1.8 TeV. About $120\,\mathrm{pb}^{-1}$ of integrated luminosity was available. The Run I remains famous due to the discovery of the top quark [44, 45] by both the DØ and CDF experiments. The second phase began in 2001 and proposed significant improvements which allowed the center-of-mass energy to increase to 1.96 TeV. The Run II is divided in two periods: Run IIa and Run IIb. The integrated luminosity at the end of Run IIa reached a bit more than $1\,\mathrm{fb}^{-1}$, which was sufficient for the CDF experiment to detect the associated production of WZ boson pairs [46], and the production of ZZ boson pairs [47] (first evidence at a hadron collider), while the DØ experiment found the first evidence for the electroweak production of single top quarks, which enabled the first direct measurement of the $t \to Wb$ coupling [48]. Run IIb started in 2006 after further improvement provided to the Tevatron, and the DØ and CDF detectors. This phase is still running and an integrated luminosity up to $8\,\mathrm{pb}^{-1}$ is expected for the final shutdown in 2009/2010. The work described in this thesis has been performed using the Run IIa data set.

In the following, we describe a collision cycle, which consists of the production, acceleration and collision of the proton and anti-proton beams.

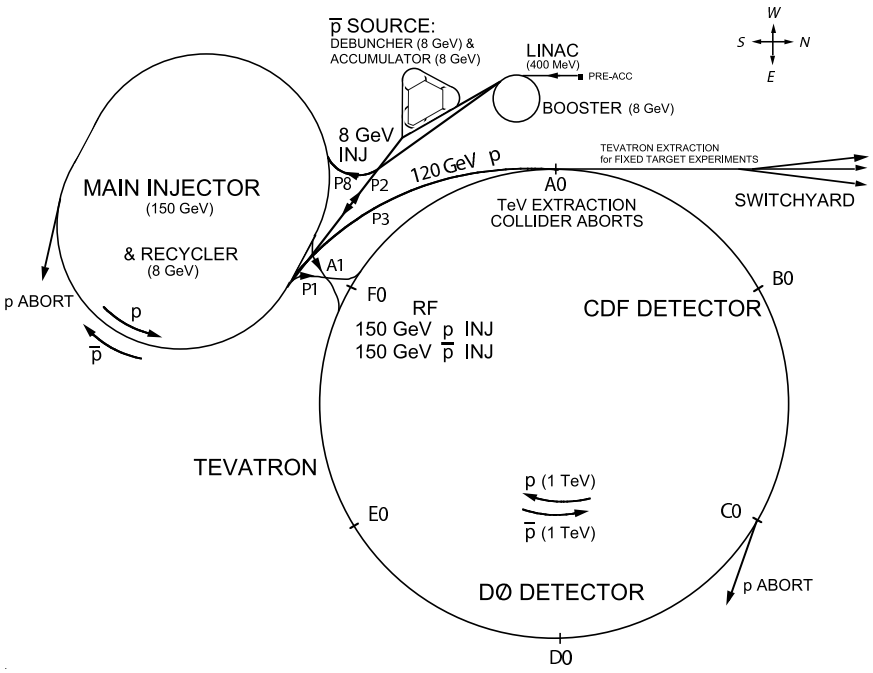

Figure 2.1: *Schematic view of the Fermilab accelerator complex.*

The proton beam

The proton beam production begins with the creation of H^- ions originating from a H_2 tank. These negative ions are passed through a Cockcroft-Walton accelerator, where they acquire an energy of 0.75 MeV. They are then injected in a linear accelerator of approximately 150 m, turning them into bunches of 400 MeV particles. In order to create a beam of protons, the hydrogen ions are passed through a carbon fiber foil, which strips off their electrons.

The first synchrotron ring of the chain, the *Booster*, accelerates the proton beam up to 8 GeV. Then, protons are passed through the *Main Injector*, a ring of about 1 km of diameter. A this stage, one part of the protons are accelerated to 150 GeV before being injected in the Tevatron, while the other part is accelerated to 120 GeV for the anti-proton production.

The anti-proton beam

The 120 GeV proton bunches are directed at a fixed target composed of nickel and copper, which produces anti-protons with a wide momenta range. About 50000 protons are necessary to create one anti-proton. The anti-protons are collected by a complex system of lithium apertures and magnetic lenses, and guided to the *debuncher* storage ring, where they are focused into a coherent 8 GeV beam. The *accumulator* storage ring further focuses the anti-protons with stochastic cooling, and distributes them in bunches to the *recycler*. The role of the *recycler* is to accumulate the anti-protons until they reach a sufficient amount to be transferred to the main injector, where they acquire an energy of 150 GeV. Since 2005, the *recycler* takes part in the cooling of the anti-protons via electron beams, which provides a better control of the anti-proton beam, and thus an increase of luminosity.

The Tevatron ring

The injection of the proton and anti-proton beams from the *Main Injector* to the Tevatron is operated when both beams have reached the energy of 150 GeV. The Tevatron ring includes about 1000 superconducting magnets working at 3.6 K and producing a magnetic field of 4.2 T. The two beams are circulating in opposite directions, and are kept separated until they both reach the maximal energy of 980 GeV. Then, the beams are collimated and meet each other in the two interaction regions BØ and DØ, where the CDF and DØ experiments are located, respectively. Each beam consists of 36 bunches, separated in time by 396 ns. The bunch crossings take place 1.7 million times each second.

2.2 The DØ detector

The DØ detector [49] is constructed of dedicated subsystems surrounding the interaction point. It has been designed to maximize the detection and identification of particles arising from $p\bar{p}$ interactions. The apparatus allows the measurement of momenta and energies, as well as the reconstruction of vertices. A general view of the DØ detector is provided in Figure 2.2.

The data streams collected by the subdetectors are filtered by triggers, depending on physics criteria, and stored to tape. At this stage, the physics objects are merely reconstructed, and are referred to as online objects. Afterwards, the reconstruction software (*DØreco*) is run offline and fully reconstructs the physics objects with more sophisticated algorithms. The offline objects used in this analysis are defined in Section 3.2.

For Run IIa, a new tracker, new calorimeter electronics, and a superconductive magnet of 2 T have been introduced. In addition, the muon system has been updated. The detector has been further improved for Run IIb, in order to be able to work under higher luminosities (new tracking layer, improved triggering system).

2.2.1 Coordinates

The Cartesian coordinate frame of the DØ detector is chosen so that its origin is placed at the center of the detector, which coincides with the nominal interaction point. The z axis has the same direction and orientation as the proton beam. The x axis is horizontal and points radially to the outside of the ring. The y axis is vertical and oriented upward.

The polar angle θ is defined in the plane z0y so that the direction $\theta = 0$ corresponds to the z axis. The azimuthal angle ϕ is defined in the transverse plan with respect to the direction of the incoming particles. The pseudo-rapidity η is preferred to the polar angle θ, is defined by the following relation:

$$\eta = -\ \ln(\tan\frac{\theta}{2}) \tag{2.1}$$

The pseudo-rapidity approximates the rapidity when the energy of the studied particle is much larger than its mass (the difference of rapidities is Lorentz invariant). $\eta_{detector}$ refers to the pseudo-rapidity defined with the center of the detector as origin, while η, also named η_{phys}, has for origin the primary vertex interaction point.

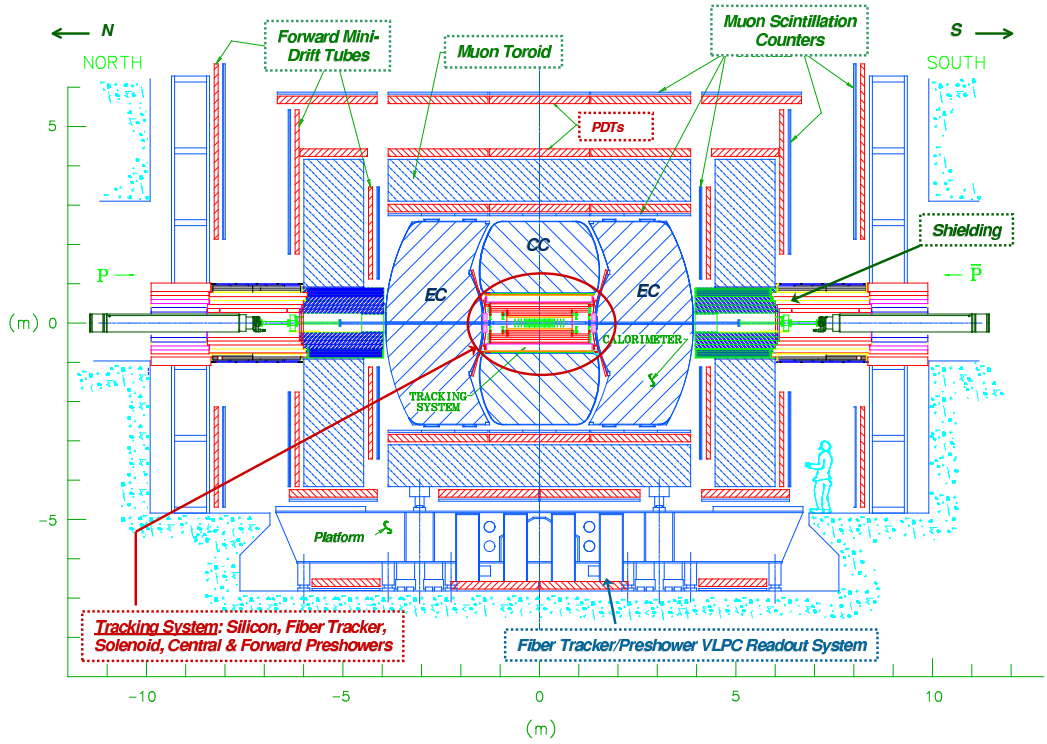

Figure 2.2: *Global side view of the DØ detector (as configured for Run II).*

The angular distance ΔR between two directions defined in η and ϕ is defined as follows:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.2}$$

Because of the structure of the proton, the colliding partons carry only a fraction of the proton kinetic energy, which implies that the center-of-mass of the parton collisions is boosted in the beam direction by an unknown amount. Quantities defined in the transverse plane, orthogonal to the beam direction, are therefore commonly used to characterize the momentum of particles. In addition, for a given collision, the sum of the transverse momenta vanishes.

2.2.2 The tracking system

The tracking system is composed of the Silicon Microstrip Tracker (SMT), which directly surrounds the interaction point, and the Central Fiber tracker (CFT), in which the SMT is embedded. Both detectors are located inside a superconductive solenoid which produces a magnetic field of 2 T, as represented in Figure 2.3.

The combination of the SMT and CFT detectors allows the reconstruction of the charged particles trajectories, which are curved by the magnetic field. The transverse momentum (p_T) can be extracted from the measurement of this curvature, with a resolution $\frac{\Delta p_T}{p_T} = 0.02 + 0.002 p_T$ at $\eta = 0$ [50].

The tracking system can be used in combination with the calorimeter to discriminate photons from electrons. As well, its association with the muon detector provides a better resolution on the muon transverse momentum.

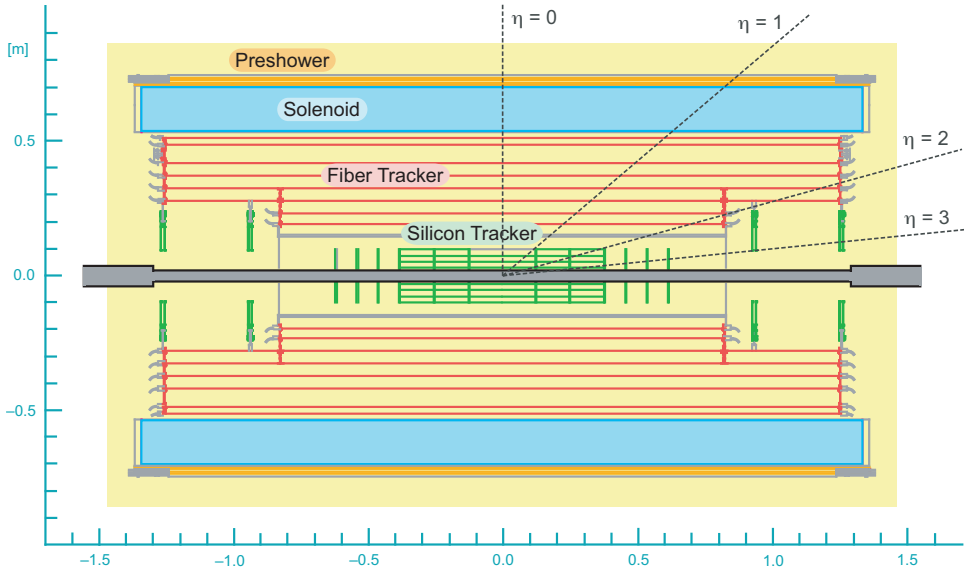

Figure 2.3: *Schematic view of the tracking system, surrounded by the superconductive solenoid magnet (2 T).*

The Silicon Microstrip Tracker detector

The SMT detector consists of wafers of silicon placed at the innermost part of the DØ detector, and allows the reconstruction of primary vertices. The ionization of the silicon by the passage of charged particles produces charges that are collected and utilized for the measurement of the particles positions.

The geometry of the SMT detector has been chosen to both cover the large distribution in z of the primary vertices and the $\eta_{detector}$ acceptance of the DØ detector. As represented in Figure 2.4, the SMT detector, centered in z, combines 6 barrels for the coverage of the region defined by $|\eta_{detector}| < 1.7$, and 16 disks for the detection of tracks up to $|\eta_{detector}| = 3$ (close to the beam axis). Each barrel is composed of 4 layers. The read-out electronics and the cooling system are located between each layer. 12 detectors of triangular shape form a disk, with a read-out system located at the periphery.

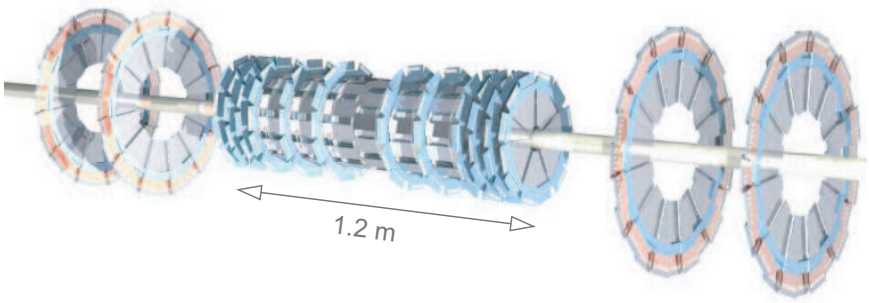

Figure 2.4: *Schematic view of the silicon miscrostrip tracker.*

The SMT detector embodies a total of 790000 read-out channels. It has been designed to

resist a radiation dose of about 1 Mrad, and is subject to 0.5 Mrad for each femtobarn of integrated luminosity. For Run IIb, a new detector layer has been installed, in order to improve the measurements, and withstand the degradations due to radiation. Currently, the SMT detector is operating fine beyond design dose.

The Central Fibre Tracking detector

The CFT detector is composed of eight concentric cylinders separated from the z axis by 20 to 51 cm. Each cylinder is covered by a layer of fibers parallel to the z axis and an additional layer of fibers which is tilted by an angle of $\pm 2^\circ$ in order to provide stereo information of the tracks along z. A total of about 77000 scintillating fibers ($\oslash \simeq 0.8$ mm) are employed.

The detection principle relies on the scintillation arising from the passage of charged particles through the fibers. The photons from scintillation are traveling along the fibers and are collected by VLPC photodetectors (Visible Light Photon Counter) to which the fibers are optically connected. The VLPC are placed outside the detector.

2.2.3 The calorimetry system

The calorimetry system surrounds the central track detectors, and consists of three main parts: the central and forward preshowers (CPS and FPS), the central and end calorimeters (CC and EC), which are embedded in their own cryostat, and the intercryostat detector (ICD).

The DØ calorimetry system has a fine granularity, and an excellent coverage and uniformity. It allows energy measurements.

The preshower (CPS and FPS)

The preshower detectors are placed between the solenoid and the cryostats. The central part (CPS) covers the region $|\eta_{detector}| < 1.3$ and the front parts (FPS) cover together the region $1.5 < \eta_{detector} < 2.5$. The goal of the preshower detectors is to provide a correction for the energy loss of particles in the solenoid and the cryostats. They also improve the distinction between pions, electron and photon showers, and allow to match the tracks to the energy deposit in the calorimeter.

The preshower detectors consist of lead absorbers and multiple scintillator strips. The readout system is similar to the one of the CFT detector.

The Central and End Calorimeters (CC and EC))

The central and end calorimeters (north and south) are represented in Figure 2.5. Each of them is surrounded by its own cryostat which allows an operating temperature of 78 K, as required for the use of liquid argon. The EC covers the region defined by $|\eta_{detector}| < 1.1$, and the EC cover the regions $\eta_{detector} < -1.5$ and $\eta_{detector} > 1.5$.

The EC and CC include four electromagnetic layers, embedded in a hadronic part, itself composed of three fine and one coarse layer. Each layer is arranged in towers composed of cells which share the same $\eta_{detector}$ and ϕ, and which have a granularity in $\eta_{detector} \times \phi$ of 0.05×0.05 in the finest electromagnetic layer (the third one from the z axis, where the maximum deposit of energy was expected), of 0.2×0.2 in the end calorimeters, and of 0.1×0.1 elsewhere.

The cells use a sampling approach, and consist of successive layers of uranium absorbers, which induce the formation of particle showers, and of active liquid argon, which is ionized by the charged particles of the arising shower.

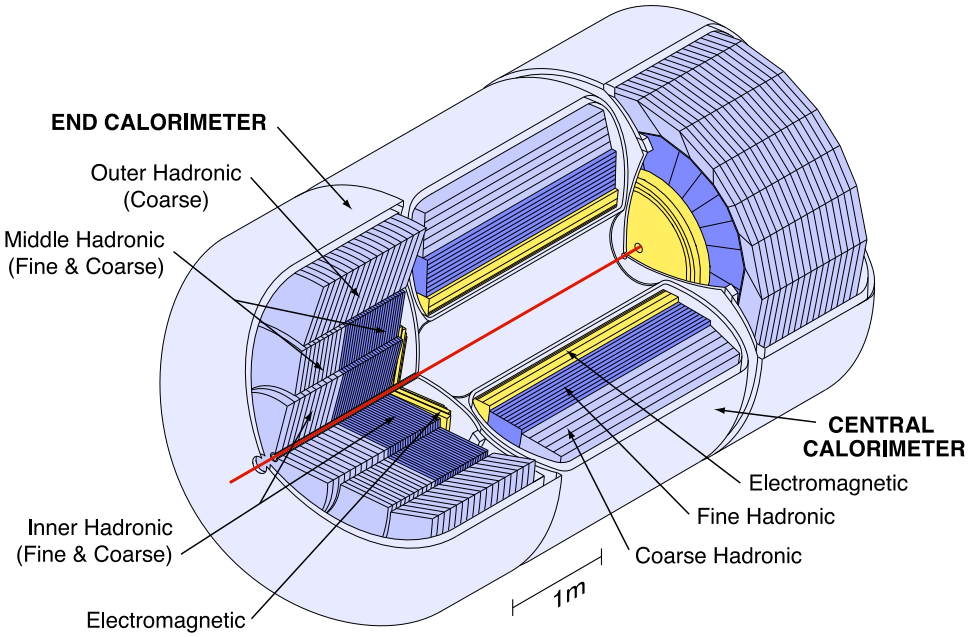

Figure 2.5: *Schematic view of the calorimeter.*

Cells of the EC and CC calorimeters which contain liquid argon but no absorber are referred to as "masseless gap" cells. They are located between the end of the "regular" calorimeter cells and the ICD. Additional coverage is thus provided for the region $0.8 < \eta_{detector} < 1.2$ of the CC, and the region $1.0 < \eta_{detector} < 1.3$.

The expected energy resolution of the calorimeter, as measured from test beam data is (with E in GeV):

$$\frac{\sigma(E)}{E} = \frac{15\%}{\sqrt{E}} + 0.3\% \text{ for electrons} \tag{2.3}$$

$$\frac{\sigma(E)}{E} = \frac{45\%}{\sqrt{E}} + 4\% \text{ for pions} \tag{2.4}$$

The Inter-Cryostat Detector (ICD)

The goal of the ICD is to overcome the lack of detectors in the region defined by $1.1 < \eta_{detector} < 1.4$. It consists of 16 layers of scintillators, which are read out via photomultipliers.

2.2.4 The muon system

Since muons are passing through the calorimeter by loosing only around 2.5 GeV, the system dedicated to their detection [51] has been placed at the periphery of the DØ detector. It consists of one central and two forward detectors, each of them being divided in three layers (A, B, and

C). Toroidal magnets, located between the layer A and the layer B, complete the system, and allow a momentum measurement. They contribute by 65% to the total mass of the detector (5500 tons). The muon chambers are arranged in planes, which gives the muon system a cubical shape. Additional iron shielding surrounds the beam pipe outside the calorimeter, so that beam remnant activity in the forward muon detectors is reduced. A schematic profile of the muon system is provided in Figure 2.6. A more global view is included in Figure 2.2.

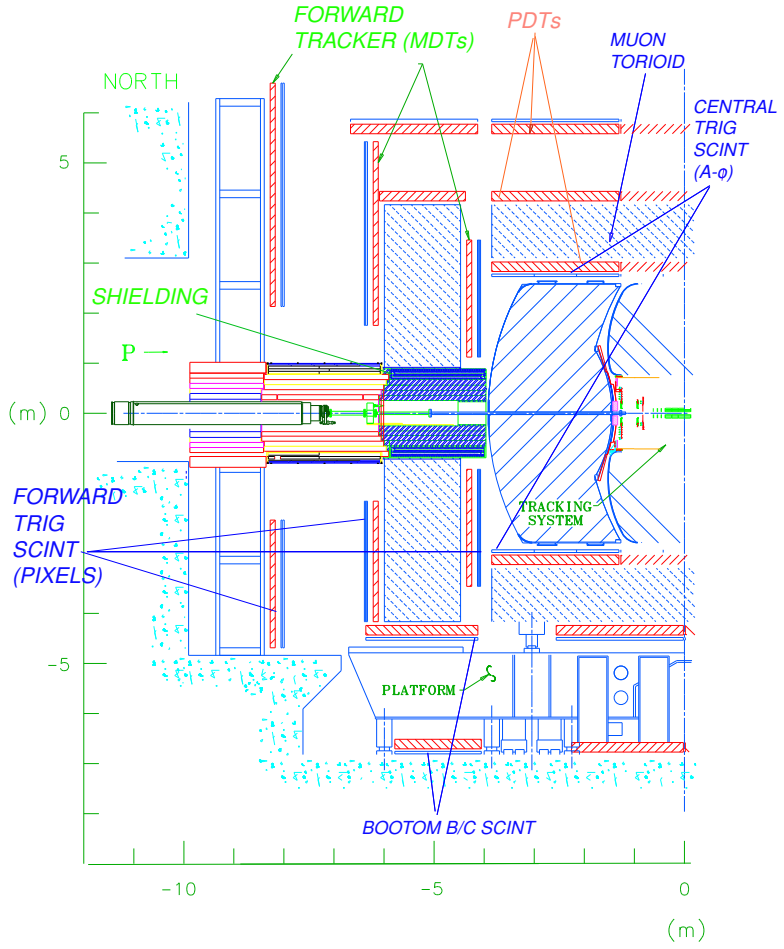

Figure 2.6: *Schematic view of the muon system.*

The muon transverse momentum resolution can be parametrized by $\frac{\sigma(p_T)}{p_T} = 0.18 + 0.03\,p_T$ for high momentum, and is limited by the individual hit resolution. For muons within the coverage of the CFT detector, the momentum measurement of the tracking detector is much better and can be associated to the muon given that the track is matched to the muon position.

The central muon detector

The central muon detector embodies 94 proportional drift tubes (PDTs) which cover the region $|\eta_{detector}| < 1$, and which are placed under a magnetic field of 1.66 T induced by the central muon toroid. The PDTs allow position measurements by collecting the charges produced by the ionization of a gas due to the passage of a particle. The A layer consists of four planes of drift

tubes, at the top of the detector, and 3 planes at the bottom. The B and C layers each contain three planes. As the drift time of the PDTs (750 ns) is longer than the bunch crossing time of the Tevatron, they cannot manage the triggering of the muons.

Layers of scintillators, the *Cosmic Cap*, are disposed on the top and the bottom of the muon detector. These 372 counters allow the rejection of cosmic muons. 630 other scintillators, the A-ϕ layers, are located between the calorimeter and the A layer drift chambers. They have been added for Run II, and allow cosmic muon rejection, improve the measurement of the muon ϕ position, and contribute to the muon triggering.

The region defined by $|\eta_{detector}| < 1.1$ and $4.75 < \phi < 5.25$, below the calorimeter, corresponds to where the support of the detector is located, as well as read-out electronics. Thus, only a few drift chambers and scintillators are covering this gap, and no space is available for any A-layer detector. This region is referred to as the *bottom hole*.

The forward muon detector

The forward muon detectors extend the muon system coverage up to $|\eta_{detector}| < 2$. In this region, the forward muon toroid induces a magnetic field of up to 1.9 T. The three layers of the forward muon detectors are composed of mini-drift tubes (MDTs) and scintillators.

The A layers consists of four planes of mini-drift tubes, arranged in octants containing 300 to 400 tubes, while the B and C layers include only three of such planes.

The scintillators are arranged in pixel arrays, and cover each of the three layers, for a total of 4214 counters. They contribute to the triggering system, allow to veto cosmic muons, and improve the direction measurement along the drift wires.

2.2.5 The triggering system

Due to the high rate of $p\bar{p}$ collisions (396 ns), the DØ detector is not able to read out and write information for each occurring interaction. The triggering system enables the selection of the collisions to be stored, and is thus one of the major component of the DØ experiment. The system must be able to both manage event selection based on physics objects, and allow fast decisions in order to write to tape a maximum of 50 events per second.

The DØ triggering system is divided in 3 levels of decision (L1, L2, L3) that are combined to define a trigger, as represented in Figure 2.7.

- The first level is purely hardware and is designed for a fast read-out. A decision is made on each bunch crossing within 4.2 μs. The selection criteria are therefore based on approximate measurements performed by each subdetectors separately, excepted the silicon tracker. At level 1, the input rate of 2.5 MHz corresponding to the beam crossing is reduced to 1.4 kHz.

- The second level decision time is lower than 100 μs. It utilizes a global processor which manages preprocessors specific to each subdetector. In other words, the information from the various subdetectors (including the SMT) are correlated, which make it possible to trigger on approximate physics objects. At this stage, the output rate is further reduced to values between 800 and 1000 Hz.

- The third level is purely software, and relies on a farm of processors. Simple reconstruction algorithms are processed within 50 ms. This level can be seen as a short version of the offline reconstruction software, to which events can be passed once stored on tape.

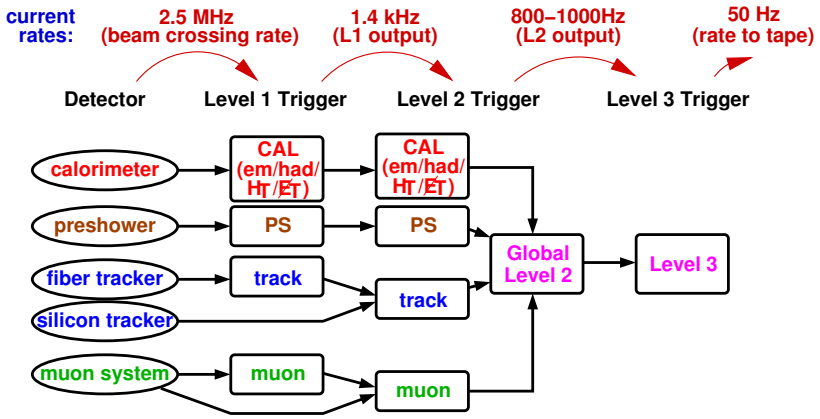

Figure 2.7: *Schematic view of the trigger framework.*

Mostly because of bandwidth limitations, prescales are applied to high-rate triggers in order to keep them below a threshold rate. Only a specified fraction of the events accepted by a prescaled trigger is actually recorded. Typically, prescales are applied for runs of the beginning of a collision cycle, when the collision rate is large. At the end of the cycle, since less collisions are occurring, prescales are likely to be loosened.

2.2.6 The luminosity measurement

The measurement of the luminosity is derived from the hits in the luminosity monitors, which signal the presence of potential inelastic $p\bar{p}$ collisions. The luminosity monitors consist of two arrays of 24 plastic scintillators which are located in the inside of the end calorimeters, in the region defined by $2.7 < \eta_{detector} < 4.4$, as described in Figure 2.8.

The luminosity is defined by the following equation:

$$\mathcal{N} = \sigma_{eff} \times \int \mathcal{L}\, dt \tag{2.5}$$

with
- $\mathcal{N}$: the number of events detected by the luminosity counters
- σ_{eff} : the effective inelastic cross section of the $p\bar{p}$ collision process, as seen by the luminosity monitors, i.e., in the forward region
- $\mathcal{L}$: the instantaneous luminosity

Figure 2.9 shows the evolution in time of the total integrated luminosity delivered by the Tevatron, and recorded by DØ. Between April 2002 and March 2008, $3.22\,\text{fb}^{-1}$ of data have been collected by DØ. Yet, the analysis described in this thesis relies on the first inverse femtobarn.

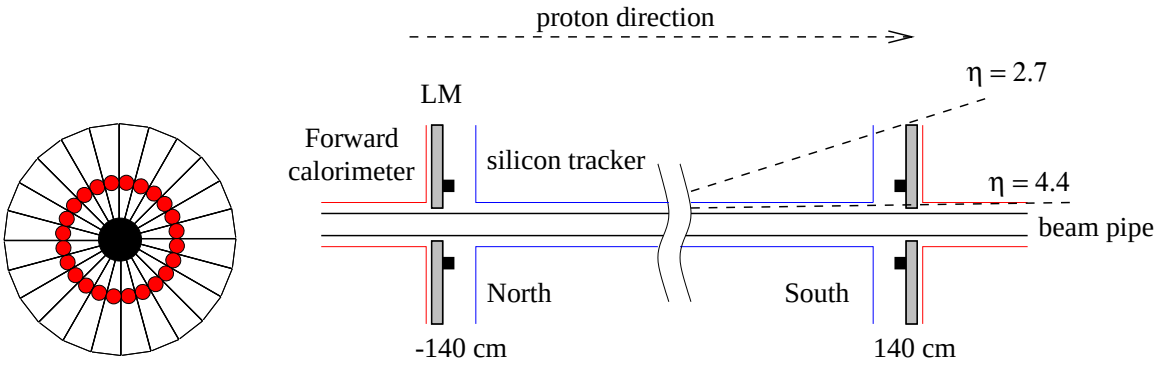

Figure 2.8: *Views of the luminosity monitors in the (x,y) (left) and the (y,z) (right) planes.*

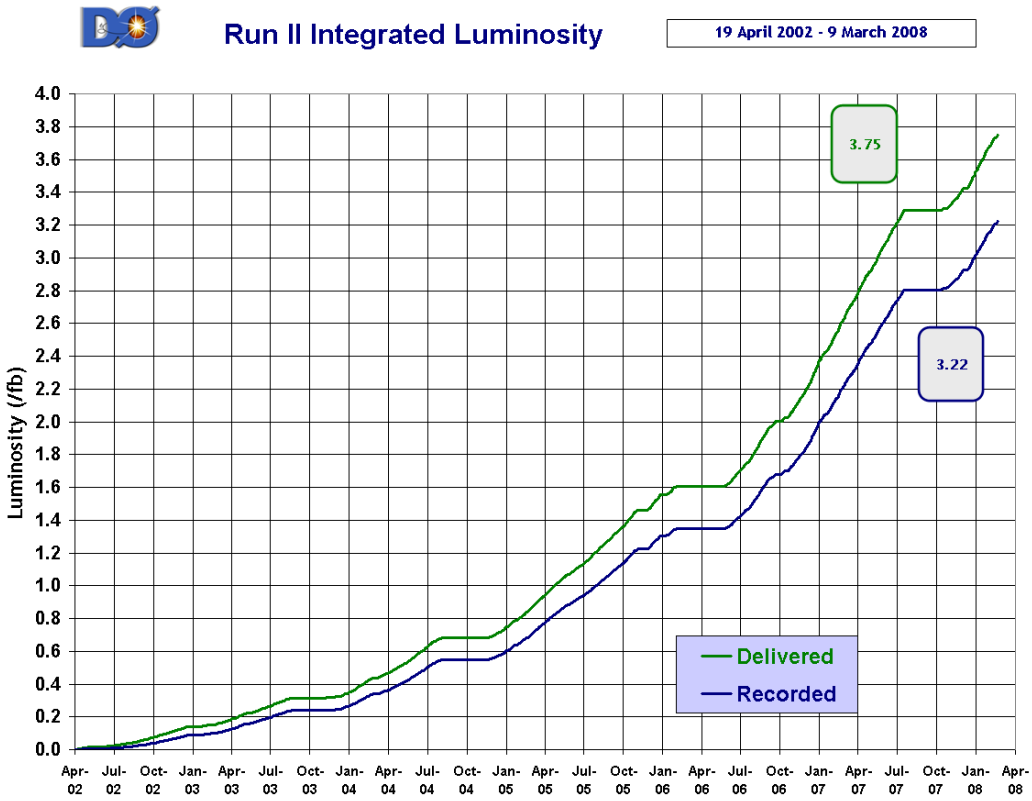

Figure 2.9: *Evolution in time of the total integrated luminosity delivered by the Tevatron, and recorded by DØ, between April 2002 and March 2008.*

Chapter 3

Data samples and physics objects reconstruction

In this chapter, the samples used for the leptoquark search are described. The definitions of the reconstructed physical objects which compose the analyzed final states are provided, and reconstruction efficiencies are detailed.

3.1 Data samples

3.1.1 The DØ data samples

The analysis presented is based on $1\,\mathrm{fb}^{-1}$ of data collected at DØ during Run IIa. which corresponds to the time period between August 2002 and February 2006, the run range [151817, 215670], and the trigger list range [v8, v14.93]. In order to perform the search for both the $LQ\overline{LQ} \to \mu\nu qq$ and $LQ\overline{LQ} \to \mu\mu qq$ decay modes, two initial data samples have been considered. One of them contains at least one muon with loose quality requirements in association with jets, and is selected with a combination of 23 single muon triggers, while the other one includes at least two muons of high transverse momenta, and is selected with a collection of 8 single muon triggers. A detailed description of the data selection is provided in Chapter 4.

Both analysis samples are required to fulfill several data quality criteria. Events included in periods of data taking where significant parts of the detector were not working properly are rejected. Typical hardware failures in the subdetectors consist of erratic read-outs, shifts in the calibration, or electronic noise. In addition, duplicated events are discarded.

3.1.2 Simulated events

The simulation chain consists in the generation of Monte-Carlo events, and the emulation of the DØ detector behavior, in order to recreate the data taking conditions.

As the first step of the detector simulation, the interactions between the generated particles and the material of the detector, as well as the effects due to the magnetic fields, are taken into account. This task is performed by the *DØgstar* software, adapted from *Geant* [52], which determines how much energy is deposited in the active areas of the detector.

Then, the detector response is simulated. The emulated interactions between the generated events and the detector components are digitized. The electronic simulation is achieved by the

software *DØsim*.

The last step is the superposition of zero-bias events, which are defined as events triggered on the unique condition that a real bunch crossing occurred. The inclusion of zero-bias events, which contain real noise, provides a realistic simulation. The data quality criteria required for real data are also applied to the zero-bias overlay.

It is worth mentioning that the detector simulation does not account for the performance of the triggering system, which must therefore be measured in data so as to be able to correct the simulated event rate accordingly. Chapter 4 describes the method used to select data events and evaluate the corresponding trigger efficiency.

Event generators

Leptoquark signals and background events arising from Standard Model processes are predicted by Monte-Carlo (MC) event generators. For the leptoquark search presented in this thesis, ALPGEN [53] and PYTHIA [54] have been used. ALPGEN is expected to provide a better jet transverse momentum shape and a better description of the jets multiplicity, as multiparton final states are explicitly included in the matrix element calculation.

- PYTHIA is based on leading order (LO) matrix elements for the generation of Standard Model and new physics processes. PYTHIA is also able to simulate parton shower, hadronization, fragmentation, and underlying events (which consists of beam-beam remnants).

  The version 6.323 with CTEQ6L1 PDF sets [30] has been used for the generation of leptoquark signal events, and for the production of $t\bar{t}$ inclusive events.

- ALPGEN is based on leading order matrix elements for X plus n partons final states. ALPGEN has been interfaced with PYTHIA for the simulation of parton shower, hadronization, fragmentation, and underlying events. Samples for various parton multiplicities are combined using the MLM matching prescription [53]. For the exclusive multiplicities, each parton produced by ALPGEN is required to be matched to exactly one jet. Only jets with a minimal transverse momentum of 8 GeV are considered, and each parton are required to be separated by a distance ΔR of at least 0.4.

  The version 2.05 with CTEQ6L1 PDF sets has been used for the W boson inclusive production in association with jets, and for the $Z/\gamma^* \to \mu\mu$ and $Z/\gamma^* \to \tau\tau$ productions in association with jets.

Luminosity reweighting

The luminosity profile in simulated events is based on the information provided by the zero-bias overlay. However, this profile might differ from the luminosity profile in the analyzed data, which is highlighted by Figure 3.1. This is due to the fact that the epoch when the zero-bias events have been collected can be different from the period of data taking corresponding to the analyzed data. In order to address the luminosity dependent effects in simulation, the luminosity profile in simulated events is reweighted so as to become as much as possible similar to the real profile in data. The luminosity reweighting does not affect the normalization of simulated samples. The event weight relative to this procedure is referred to as $\mathcal{W}_{\mathcal{L}}$.

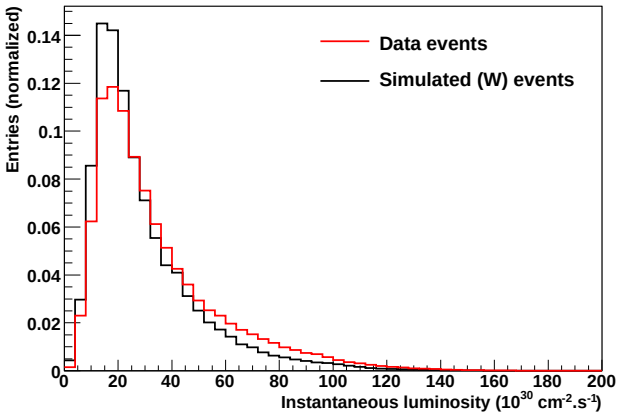

Figure 3.1: *Comparison of the luminosity profile in data and in simulated (*PYTHIA*) W boson production.*

Simulated samples

The same background from the Standard Model processes is considered in both analyzed channels. The QCD multijet contribution has been estimated from the data sample for the $LQ\overline{LQ} \to \mu\nu jj$ analysis, as described in Section 5.1.1. For the $LQ\overline{LQ} \to \mu\mu jj$ study, the QCD multijet contribution can be neglected with an appropriate event preselection, which is detailed in Section 5.2.1.

The list and the production properties of all the samples that have been used to simulate the Standard Model background processes are provided in Table 3.1. W boson production in association with jets has been simulated with ALPGEN and includes the three leptonic decays. Samples of $Z/\gamma^* \to \mu\mu$ and $Z/\gamma^* \to \tau\tau$ have been generated with ALPGEN for the di-lepton mass range from 15 GeV to 1960 GeV (kinematic limit of the Tevatron). The $t\bar{t}$ inclusive production has been simulated by PYTHIA. Next-to-next-to leading order cross sections have been considered for the W and Z/γ^* samples, while NLO predictions have been used for the $t\bar{t}$ background. The Z/γ^* samples have been scaled by a k-factor depending on the di-muon or di-tau invariant mass.

The signal samples have been produced for ten different leptoquark masses: 140, 160, 180, 200, 220, 240, 260, 280, 300 and 320 GeV. From 40 to 50k events have been generated for each mass. The cross sections were calculated at next-to-leading order (NLO) [29], and vary from 0.0739 to 2.38 pb depending on the assumed mass, as listed in Table 3.2. For both decay modes considered, we forced the leptoquarks to only decay to second generation quarks.

The equivalent luminosities of simulated samples have been determined after removing duplicated events, as well as events qualified as bad according to the data quality requirements. In addition, events which do not contain the instantaneous luminosity information have also been discarded, since this information is needed to perform the luminosity reweighting. The number of events in Tables 3.1 and 3.2 therefore refers to the number of simulated events remaining after imposing the preceding requirements.

Sample		$\sigma \times Br$ (pb)	#Events	MC Generator
$W(+jets) \to \ell\nu + jets$ ($\ell = e, \mu, \tau$)		7748 (NNLO)	26007k	ALPGEN+PYTHIA
$Z/\gamma^*(+jets) \to \mu\mu + jets$	$\frac{M(\mu,\mu)}{\text{GeV}} \in$ [15-60]	386.7 (LO)	1366k	ALPGEN+PYTHIA
	$\frac{M(\mu,\mu)}{\text{GeV}} \in$ [60-130]	192.6 (LO)	3288k	
	$\frac{M(\mu,\mu)}{\text{GeV}} \in$ [130-250]	1.38 (LO)	966k	
	$\frac{M(\mu,\mu)}{\text{GeV}} \in$ [250-1960]	0.125 (LO)	376k	
$Z/\gamma^*(+jets) \to \tau\tau + jets$	$\frac{M(\tau,\tau)}{\text{GeV}} \in$ [15-60]	388.3 (LO)	3555k	ALPGEN+PYTHIA
	$\frac{M(\tau,\tau)}{\text{GeV}} \in$ [60-130]	195.4 (LO)	1284k	
	$\frac{M(\tau,\tau)}{\text{GeV}} \in$ [130-250]	1.42 (LO)	373k	
	$\frac{M(\tau,\tau)}{\text{GeV}} \in$ [250-1960]	0.129 (LO)	374k	
$t\bar{t}$ inclusive		6.77 (NLO)	1550k	PYTHIA

Table 3.1: *List of all the simulated samples that have been used to describe the backgrounds. $M(\mu,\mu)$ and $M(\tau,\tau)$ refers to the di-muon and di-tau invariant masses, respectively. The given cross sections for the W sample is calculated at next-to-next-to-leading order (NNLO) [55], and the one for the $t\bar{t}$ sample is computed at next-to-leading order (NLO) [56]. The listed cross sections for the Z/γ^* samples are determined at LO by* ALPGEN. *For $M(\mu,\mu) \in$[60-130] GeV, the NNLO cross section for the $Z/\gamma^* \to \mu\mu$ process is 256.6 pb [55].*

M_{LQ} (GeV)	σ (pb)	#Events	MC Generator
140	2.38	from about 40k to 50k each	PYTHIA
160	1.08		
180	0.525		
200	0.268		
220	0.141		
240	0.0762		
260	0.0419		
280	0.0233		
300	0.0131		
320	0.00739		

Table 3.2: *List of all the simulated signal samples. The cross sections are calculated at NLO [29].*

3.2 Physics objects reconstruction

In the case of the search in the $LQ\overline{LQ} \rightarrow \mu\nu qq$ decay mode, one of the leptoquarks is assumed to decay to a muon and a quark, and the other one to a neutrino and a quark, which leads to a final state composed of a muon, missing transverse energy and two jets ($\mu \not\!\!E_T jj$ signature). By contrast, in the case of the search in the $LQ\overline{LQ} \rightarrow \mu\mu qq$ decay mode, both leptoquarks are assumed to decay to a muon and a quark, which leads to an event topology consisting of two muons and two jets ($\mu\mu jj$ signature). As a result, the physics objects referred to as muon, jet and missing transverse energy ($\not\!\!E_T$) have to be identified. For both real data and simulated events, the reconstruction of these physics objects, which we briefly describe in the following, is led by the same software, namely *DØreco*.

3.2.1 Jet reconstruction

Jets are physical objects consisting of the particles arising from the hadronization of a quark or a gluon. Jets are subject to interactions in the calorimeter, where they deposit energy.

Reconstruction algorithm

In this analysis, jets are reconstructed with the Run II cone algorithm [57] using a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.5$. Towers of the calorimeter containing an energy of at least 0.5 GeV are first considered as preliminary jets. Then, the direction of each jet is calculated depending on its spatial energy distribution, by taking the primary vertex as origin. When two jets are separated by an angular distance lower than $2\Delta R$, their common energy is calculated. If this energy is larger than a defined threshold, the two jets are merged into one, otherwise, they remain separated and the more energetic one absorbs the common energy. Additional iterations (direction calculation and eventual merging) are performed until all the jets have no common energy. Finally, only jets with energies greater than 6 GeV are kept.

Jet identification

Each selected jet must fulfill standard quality criteria [58] which include requirements on the electromagnetic and coarse hadronic fractions of the cell energies and a L1 trigger confirmation.

- Electromagnetic fraction:

  The fraction of energy deposited by jets in the electromagnetic calorimeter (EMF) must fulfill at least one of the following criterion:

  $$\begin{cases} 0.05 < EMF \\ 0.03 < EMF \text{ and } 1.1 < |\eta| < 1.4 \\ 0.04 < EMF \text{ and } 2.5 < |\eta| \end{cases}$$

  The jet must also verify $||\eta_{detector}| - 1.25| + 0.40 \times (\sigma_n - 0.1) < 1.3$, where σ_n is the maximal η width of the jet. In this region, the calorimeter is weakly equipped with electromagnetic layers.

- Coarse hadronic fraction:

  The fraction of energy deposited by jets in the coarse hadronic calorimeter (CHF) is required to fulfill at least one of the following conditions:

$$\begin{cases} CHF < 0.4 \\ CHF < 0.6 \quad \text{and} \quad 0.85 < |\eta_{detector}| < 1.25 \quad \text{and} \quad n90 < 20 \\ CHF < 0.44 \quad \text{and} \quad |\eta| < 0.8 \\ CHF < 0.6 \quad \text{and} \quad 1.5 < |\eta| < 2.5 \end{cases}$$

  where $n90$ is the minimal number of towers containing 90% of the jet energy

- L1 trigger confirmation:

  Two read-out streams are provided by the calorimeter. A coarse one is used for the level 1 trigger decision, and a precise one is utilized for the reconstruction. Since the high gain of the precise stream is subject to noise which could be interpreted as physics objects, the level 1 confirmation is required. The ratio between the jet p_T as measured at L1 and the jet p_T after reconstruction is defined as $L1_{ratio}$. At least one of the following conditions must be fulfilled:

$$\begin{cases} 0.5 < L1_{ratio} \\ 0.35 < L1_{ratio} \quad \text{and} \quad p_T < 15\ \text{GeV} \quad \text{and} \quad |\eta| > 1.4 \\ 0.2 < L1_{ratio} \quad \text{and} \quad p_T > 15\ \text{GeV} \quad \text{and} \quad |\eta| > 3.0 \\ 0.1 < L1_{ratio} \quad \text{and} \quad p_T < 15\ \text{GeV} \quad \text{and} \quad |\eta| > 3.0 \end{cases}$$

Jet Energy Scale

The jet energies (either data or simulation) are calibrated as a function of the jet transverse energy and η by balancing the transverse energy in photon plus jet events. This correction is referred to as the Jet Energy Scale (JES) correction [59].

The calibration consists in relating the jet energy measured in the detector (E_{jet}^{meas}) to the final state particle jet energy (E_{jet}^{ptcl}). The following relation summarizes the effects that are taken into account:

$$E_{jet}^{ptcl} = \frac{E_{jet}^{meas} - E_O}{R_{jet}\, S_{jet}} \tag{3.1}$$

where:

- E_O is the offset energy, namely the energy in the jet cone which is not associated to hard scattering interactions (electronic and uranium noise, additional $p\bar{p}$ interactions, pile-up from previous beam crossings)
- R_{jet} is the calorimeter response, which is a function the jet energy, the jet η position, and the jet cone radius.
- S_{jet} is the jet energy fraction which has been deposited out of the cone, and thus not included in E_{jet}^{meas}

Jet shifting, smearing and removal

Simulated jets have to be further corrected in order to reach agreement with real data. Differences in data and simulation can arise from the jet energy scale [60], the energy resolution, and the reconstruction and identification efficiencies.

The resolution of the jets is determined by $\gamma + jets$ events, and is further recalibrated (shifted) according to the relative data-simulation jet energy scale. In addition, a jet p_T threshold of 15 GeV is required both in data and simulation. At a p_T of 15 GeV, the jet reconstruction and identification efficiencies have nearly reached their plateau.

The three corrections are performed consistently by comparing real and simulated $\gamma + jets$ events. Systematic errors are derived by comparing results obtained with $\gamma + jets$ and $Z \rightarrow ee + jets$ events. More details about the complete procedure can be found in Reference [61].

3.2.2 Muon reconstruction

The muons are reconstructed using hits in the layers of the muon detector. For the leptoquark search presented in this document, all the muons are required to be matched to a track reconstructed from hits in the central tracking system, in order to improve the transverse momentum resolution. A veto on cosmic muons, based on timing information in the muon system, is applied in addition. In the following, we define the quality criteria that have been utilized.

Muon identification qualities

A muon of *loose* quality does not require to be reconstructed with hits in each of the three layers. It must yet fulfill at least one of the following criteria (the layers of the muon detector are described in Section 2.2.4):

- in the A layer, at least one scintillator hit and at least two hits in the drift tubes
- in the BC layer, at least one scintillator hit and at least two hits in the drift tubes
- at most one of the following criterion fails:
  - in the A layer, at least two hits in drift tubes and at least one scintillator hits
  - at least two hits in the BC layer drift tubes
  - at least one scintillator hit in the BC layer

A muon of *mediumnseg3* quality is reconstructed with hits in the three layers fulfills all the following criteria:

- at least two hits in the A layer drift tubes
- at least one scintillator hits in the A layer
- at least two hits in the BC layer drift tubes
- at least one scintillator hit in the BC layer

Matched track and transverse momentum oversmearing

Each matched central track must have a distance of closest approach (*dca*) to the primary vertex smaller than 0.2 cm (in this case the *dca* is the minimal distance between the track and the primary vertex position). In addition, the χ^2 per degree of freedom of the central track fit is required to be smaller than 4. A track fulfilling these conditions is referred to as a *medium* track.

The matched central track is used to measure the value of the transverse momentum of the muon, which is further smeared in simulated events so that the width of the $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ peaks match that observed in data [62]. The smearing is performed via the following substitution:

$$\frac{q}{p_T} \to \frac{q}{p_T} + A\,G_1 + \frac{B\,\sqrt{cosh(\eta)}}{p_T}\,G_2 \tag{3.2}$$

where q is the charge of the muon, p_T is the transverse momentum of the muon, and G_1 and G_2 are two independent random numbers distributed according to the normal law.

Depending on the run number and the pseudo-rapidity η as measured in the CFT detector, the number A varies between $1.7 \cdot 10^{-3}$ and $3.1 \cdot 10^{-3}\,\mathrm{GeV}^{-1}$, while the number B takes values between $0.9 \cdot 10^{-2}$ and $2.5 \cdot 10^{-2}$.

Muon isolation qualities

The matched track isolation is ensured by requiring an upper threshold on the sum of transverse momenta of all other tracks in a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.5$ around the muon This sum is referred to as $EtTrkCone5$. A muon is of *loose isolation* if $EtTrkCone5$ is below 4 GeV. A tight isolation requires $EtTrkCone5 < 2.5\,\mathrm{GeV}$.

The muon isolation can be further tightened by requiring an upper threshold on the energy deposited in the calorimeter in an annulus of radius $0.1 < \Delta R < 0.4$ around the muon. The energy deposited is referred to as $EtHalo$. The tight muon isolation criterion is completed when $EtHalo$ does not exceed 2.5 GeV.

3.2.3 Neutrinos and missing transverse energy

In each collision, the partons which do not participate in the hard process hadronize into beam-remnants due to confinement. Those remnants have a large longitudinal boost into the direction of the proton or anti-proton, respectively, and thus are outside the acceptance of the detector. The total energy of the entire final state of the scattering process is thus not measurable. As the lost particles have a very small transverse energy, the vectorial sum of the all the measured transverse energies is expected to be negligible. The missing transverse energy ($\overrightarrow{\not{E}_T}$), arising from particles that pass through the detector without being detected, can thus be defined as the opposite of the vectorial sum of all the measured transverse energies.

First the missing transverse energy is estimated as being the opposite of the vectorial sum of the positive transverse energies measured in all the cells of the calorimeter, except those of the coarse hadronic layers, since it is subject to substantial noise. The coarse hadronic cells belonging to jets are however included in the calculation.

The second step consists in propagating the calibration corrections provided to the reconstructed jets (for electrons the calibrated cluster is taken as well). No propagation of the jet energy scale is yet included for the coarse hadronic cells belonging to jets.

Since the muons are not stopped by the calorimeter, the missing transverse energy is finally corrected by removing the vectorial transverse momentum of the muons, and adding back the small amount of energy they deposit in the calorimeter (the latter is not measured, but the expectation is used).

Neutrinos do not deposit any energy in the detector, but their transverse momentum can be inferred from the missing transverse energy.

3.3 Muon reconstruction efficiencies

The muon identification, tracking, and isolation efficiencies are a priori not identical in data events and in the simulation. Each simulated event is therefore corrected to reach agreement with data. The correction factors which account for the muon identification, the tracking, and the muon isolation efficiencies are obtained by dividing the efficiencies observed in data over those obtained from simulated events.

For a given simulated event, the muon reconstruction efficiency factor is thus defined by:

$$\mathcal{W}_{\mu-reco} = \left(\frac{\varepsilon^{data}_{muon-id}}{\varepsilon^{MC}_{muon-id}}\right)^m \times \left(\frac{\varepsilon^{data}_{tracking}}{\varepsilon^{MC}_{tracking}}\right)^m \times \left(\frac{\varepsilon^{data}_{muon-iso}}{\varepsilon^{MC}_{muon-iso}}\right)^m \tag{3.3}$$

where $\varepsilon_{muon-id}$, $\varepsilon_{tracking}$, $\varepsilon_{muon-iso}$ are the muon identification, tracking, and muon isolation efficiencies, respectively. m is the number of selected muons in the given event.

All the results provided in the following have been obtained by following the official muon identification group prescriptions [63].

The tag-and-probe method

Efficiencies relative to the muon reconstruction are estimated with the tag-and-probe method. The decay of the Z boson into two muons is examined through data and simulated events (PYTHIA prediction).

The event selection requires one (tag) muon of good quality. It has to be matched to a central track, have a p_T greater than 30 GeV, and fulfill $EtHalo < 2.5$ GeV and $EtTrkCone5 < 3.5$ GeV. The Probe object can either be a central track or a local muon of good quality ($p_T > 20$ GeV, same isolation as the tag), but the invariant mass of the tag muon and the probe object must be in the mass region of the Z boson. Matching a probe central track or a probe local muon to, respectively, a local muon or a central track will provide muon detector or central tracking efficiencies. Figure 3.2 represents the tag-and-probe method as utilized for the estimation of the tracking efficiency.

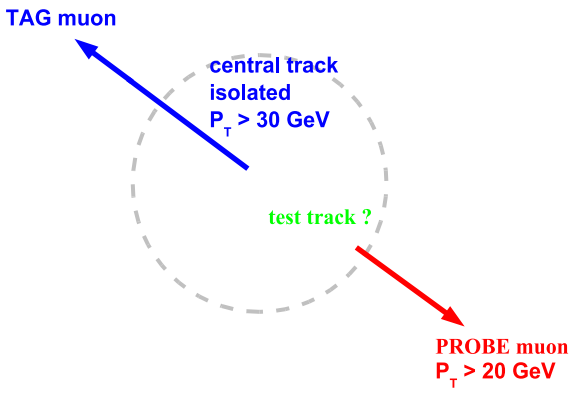

Figure 3.2: *Schematic representation of the Tag and Probe method in the case of the central tracking efficiency measurement. The tag muon and the probe muon are of good quality.*

The muon identification efficiency

The efficiencies of the track matching and of the cosmic veto are included in the muon identification efficiency, which is parametrized in bins of $\eta_{detector}$ and ϕ. The muon identification efficiency includes the track-matching efficiency, i.e., the efficiency that a track is matched to the muon.

The *loose* and *mediumnseg3* identification efficiencies, for both data and simulated events, are provided in Figures 3.3 and 3.4, respectively. It is worth mentioning that a muon, when required to be found with hits in the three layers of the muon system (as for a *mediumnseg3* muon), cannot be detected in the bottom hole of the detector due to its poor instrumentation. As a result, Figure 3.4 presents an "efficiency hole" in the (ϕ,$\eta_{detector}$) map distribution.

Table 3.3 contains the values of the muon identification efficiencies in data and simulation as calculated with the tag-and-probe method, and the correction factors $\left(\frac{\varepsilon^{data}_{muon-id}}{\varepsilon^{MC}_{muon-id}}\right)^{m}$ for the simulated samples of both channels, after the preselection cuts defined in Chapter 5. The $Z/\gamma^* \rightarrow \tau\tau$ sample is not shown in the table since negligible due to low statistics after the preselection cuts.

	Tag-and-probe results			Weights in analysis samples ($\frac{data}{MC}$)			
Quality (Channel)	data	$Z/\gamma^* \rightarrow \mu\mu$	$\frac{data}{MC}$	W	$Z/\gamma^* \rightarrow \mu\mu$	$t\bar{t}$	signal
loose ($\mu\mu jj$)	89.43%	90.52%	0.988	-	0.985	0.979	0.980
mediumnseg3 ($\mu \not{E}_T jj$)	70.08%	72.93%	0.961	0.980	0.980	0.974	0.975

Table 3.3: *Muon identification efficiencies in data and simulation and resulting weights in both search channels. Signal mass is assumed to be 260 GeV. "-" refers to the absence of value because of statistical irrelevance.*

The tracking efficiency

The tracking efficiency is parametrized in bins of track z position, and η_{CFT} (η in the CFT detector). Figure 3.5 represents the tracking efficiency for a medium track, both in the case of data and simulated events. The two plots of the Figure highlight the large acceptance in η and z

of the tracking system. Regions of low efficiencies are located at the extremities of cylinder, and correspond to muons which do not travel a large distance across the detector ($\eta \times z > 0$).

Table 3.4 provides the tracking efficiencies in data and simulation as calculated with the tag-and-probe method, and the correction factors $\left(\frac{\varepsilon^{data}_{tracking}}{\varepsilon^{MC}_{tracking}}\right)^m$ for the simulated samples of both channels, at preselection level (the $Z/\gamma^* \to \tau\tau$ sample is not considered since not statistically relevant).

	Tag-and-probe results			Weights in analysis samples ($\frac{data}{MC}$)			
Quality (Channel)	data	$Z/\gamma^* \to \mu\mu$	$\frac{data}{MC}$	W	$Z/\gamma^* \to \mu\mu$	$t\bar{t}$	signal
medium ($\mu\mu jj$)	90.19%	96.40%	0.936	-	0.869	0.852	0.847
medium ($\mu \not{E}_T jj$)				0.933	0.932	0.921	0.919

Table 3.4: *Tracking efficiencies in data and simulation and resulting weights in both search channels. Signal mass is assumed to be 260 GeV. "-" refers to the absence of value because of statistical irrelevance.*

The muon isolation efficiency

The isolation efficiency is parametrized in bins of jet multiplicity and muon p_T. The loose and tight isolation efficiencies, both for data and simulated events, are provided in Figure 3.6 and 3.7, respectively. The four plots show a decrease of the efficiency with increasing number of jets, which is explained by the fact that muons are more likely to be close to jets in multijet topologies.

Table 3.5 lists the muon isolation efficiencies in data and simulation as calculated with the tag-and-probe method, and the correction factors $\left(\frac{\varepsilon^{data}_{muon-iso}}{\varepsilon^{MC}_{muon-iso}}\right)^m$ for the simulated samples of both channels, at preselection level (the $Z/\gamma^* \to \tau\tau$ sample is not considered since not statistically relevant).

	Tag-and-probe results			Weights in analysis samples ($\frac{data}{MC}$)			
Quality (Channel)	data	$Z/\gamma^* \to \mu\mu$	$\frac{data}{MC}$	W	$Z/\gamma^* \to \mu\mu$	$t\bar{t}$	signal
loose ($\mu\mu jj$)	94.67%	94.36%	1.00	-	1.04	1.06	1.19
tight ($\mu \not{E}_T jj$)	92.52%	92.49%	1.00	1.00	1.01	0.996	1.03

Table 3.5: *Muon isolation efficiencies in data and simulation, and resulting weights in both search channels. Signal mass is assumed to be 260 GeV. "-" refers to the absence of value because of statistical irrelevance.*

3.4 Event weight in simulated samples

To sum things up, all simulated samples are normalized to the data sample luminosity, and corrected with the luminosity reweighting factors ($\mathcal{W}_{\mathcal{L}}$), the trigger weights ($\mathcal{W}_{trigger}$), described in the following chapter, and the muon reconstruction weights $\mathcal{W}_{\mu-reco}$ defined in the preceding section.

For a sample of N generated events simulating a process of cross section σ, and for a data luminosity L, the global event correction factor is thus:

$$\mathcal{W}_{event}^{MC} = \mathcal{W}_{trigger} \times \mathcal{W}_{\mu-reco} \times \frac{L \times \sigma}{N} \times \mathcal{W}_{\mathcal{L}} \tag{3.4}$$

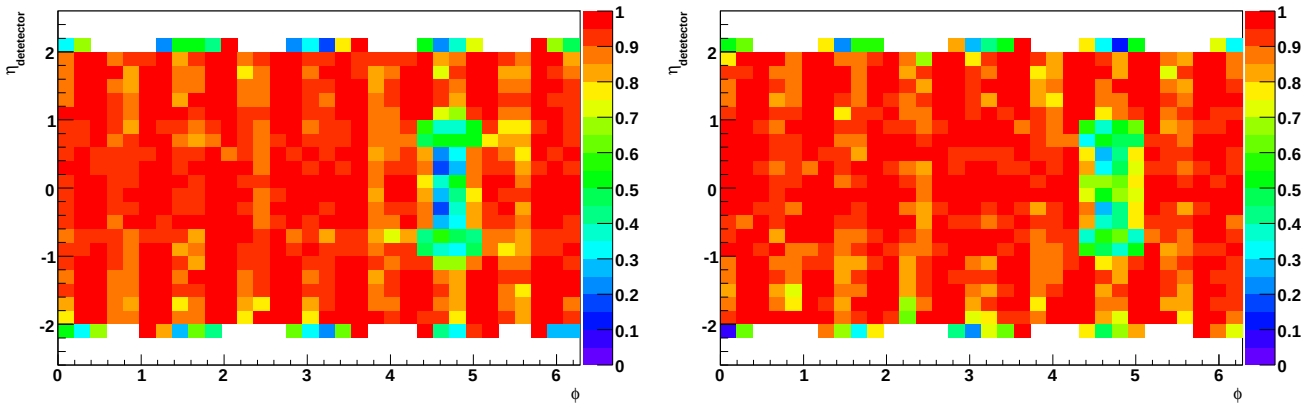

Figure 3.3: *Reconstruction efficiency for muons of loose quality, in the case of data (left) or simulated (right) events.*

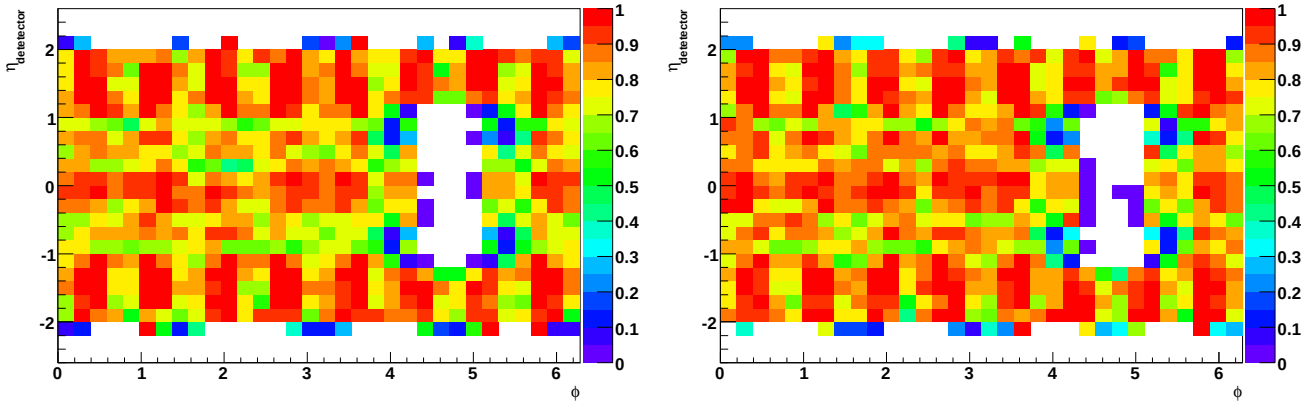

Figure 3.4: *Reconstruction efficiency for muons of mediumnseg3 quality, in the case of data (left) and simulated (right) events.*

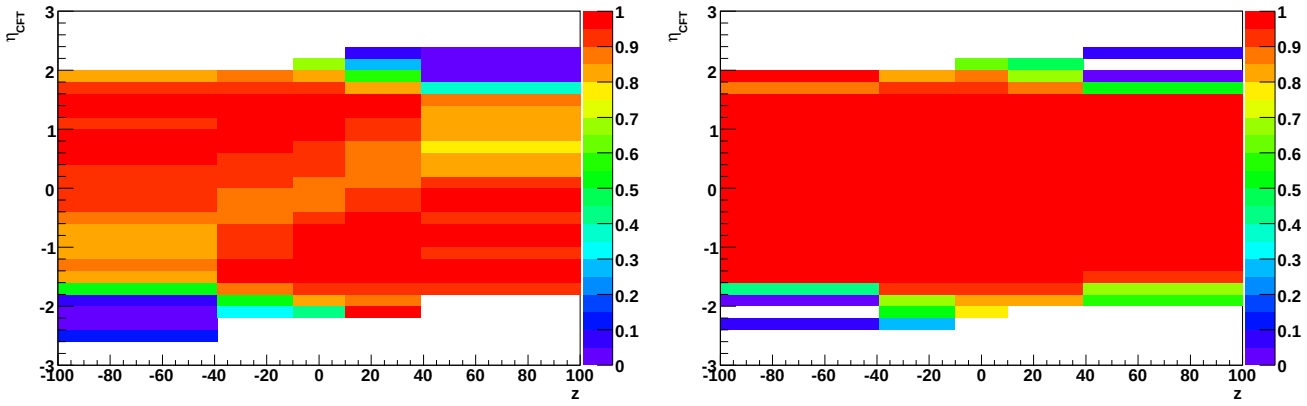

Figure 3.5: *Reconstruction efficiency for a track of medium quality, both in the case of data (left) and simulated events (right).*

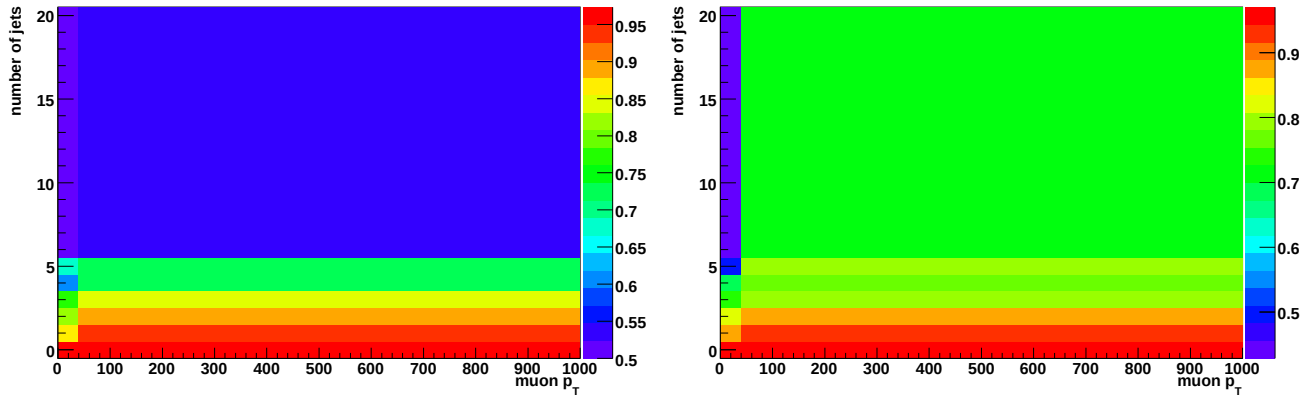

Figure 3.6: *Reconstruction efficiency for a muon of loose isolation quality, both in the case of data (left) and simulated events (right).*

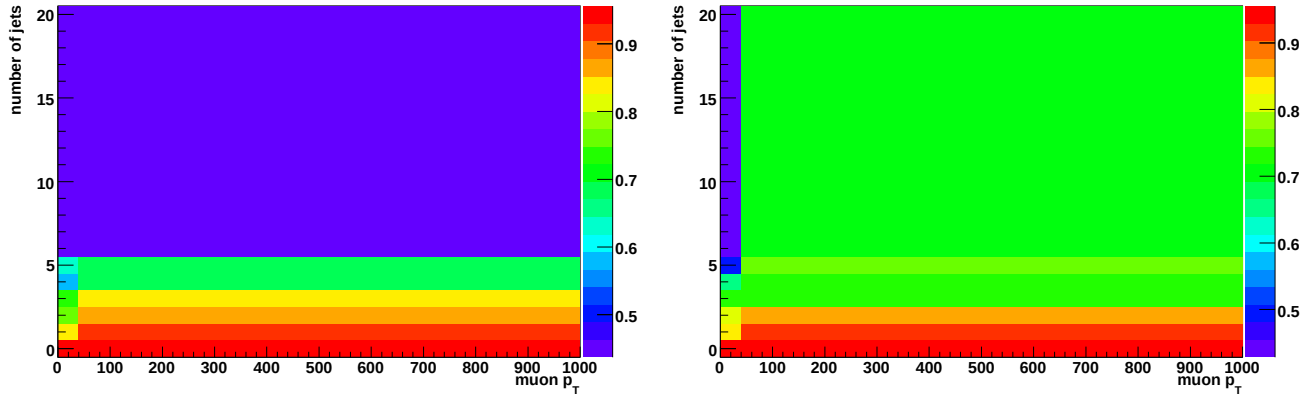

Figure 3.7: *Reconstruction efficiency for a muon of tight isolation quality, both in the case of data (left) and simulated events (right).*

Chapter 4

Combination of single muon trigger by using an inclusive OR

The analysis presented in this thesis is based on the data taken at the Fermilab Tevatron collider ($\sqrt{s} = 1.96$ TeV) during Run IIa. In order to increase the statistics available for the data sample, and thus improve the sensitivity of the leptoquark search, several triggers have been considered simultaneously for the data selection. Since both analyzed channels include finale states composed of muons, a collection of single muons triggers have been used to select the data.

For each main period of data taking, the triggers that are important for physics analysis are always kept unprescaled. However, depending on runs, additional triggers may be configured with varying prescales. Combining the main unprescaled triggers with the additional prescaled triggers thus requires to a consider a large variety of prescales.

Simulated events, which are used to describe the data, need to be reweighted in order to account for the trigger performance. Since Monte-Carlo events are also normalized to the luminosity, the prescales can be included either in the trigger efficiency estimation, or in the luminosity calculation. The former entails significant complications while the latter cannot be achieved when triggers with different prescales are combined.

This chapter presents the strategy used to combine 23 single muon triggers with an inclusive OR. The combination includes both prescaled and unprescaled triggers, and is utilized for the selection of final states containing exactly one muon ($\mu \not\!\!E_T jj$ channel). More details can be found in reference [64]. In the case of topologies containing more than one muon ($\mu\mu jj$ channel), only eight unprescaled triggers are combined, in order to avoid a possible bias due to correlations between the muons.

4.1 Single muon triggers in Run IIa

4.1.1 Trigger definitions and trigger lists

Each of the three single muon trigger levels are composed of specific conditions dedicated to muon or track detection. Table 4.1 provides the description of all the single muon triggers that have been studied for the combination. All the associated trigger conditions are defined in Appendix A.

For each run, a list of triggers is used to specify which trigger should operate and with which prescale. Unprescaled triggers have a "prescale factor" of 100%. The period of successive runs

for which the same trigger list has been utilized is referred to as the "trigger list period". The different trigger list periods are indexed by a version number, namely the "trigger list version". The period of activity given for each trigger in Table 4.1 provides the range of trigger list versions corresponding to the epoch where the trigger was configured to operate. Expressed in terms of trigger list versions, the Run IIa epoch corresponds to the trigger list range [v8.00,v14.93].

Trigger name	Level 1	Level 2	Level 3	Activity
MU_W_L2M5_TRK10	L1WTN0	L2M5	L3TRK10	8.00 - 10.03
MU_W_L2M0_TRK3	L1WTN0	L2M5	L3TRK3H10	8.10 - 9.50
MUW_W_L2M5_TRK10	L1WTL0	L2M5	L3TRK10	10.00 - 10.03
MU_W_L2M0_TRK10	L1WTN0	L2M3	L3TRK10	9.30 - 9.50
MU_W_L2M3_TRK10	L1WTN0	L2M3	L3TRK10	10.30 - 12.37
MUW_W_L2M3_TRK10	L1WTL0	L2M3	L3TRK10	10.30 - 12.37
MUW_A_L2M3_TRK10	L1ATL0	L2M3	L3TRK10	10.30 - 11.04
MUH1_TK12	L1WTN10, L1TRK10	-	L3L0, L3CML12, L3TRK12	13.10 - 13.11
MUH1_TK10	L1WTN10, L1TRK10	-	L3TRK10	13.03 - 13.62
MUH1_LM15	L1WTN10, L1TRK10	-	L3L15	13.03 - 13.90
MUH1_TK12_TLM12	L1WTN10, L1TRK10	-	L3L0, L3CML12, L3TRK12H10	13.20 - 14.93
MUH1_ILM15	L1WTN10, L1TRK10	-	L3L15, ISO_MUON_CAL3 ISO_MUON_CM_L1	14.00 - 14.93
MUH2_LM15	L1ATL0, L1TRK10	L2M3	L3L15	13.03 - 13.23
MUH2_LM10_TK12	L1ATL0, L1TRK10	L2M3	L3L10, L3TRK12	13.30 - 13.90
MUH2_LM6_TK12	L1ATL0, L1TRK10	L2M3	L3L6, L3TRK12	13.20 - 13.23
MUH2_LM3_TK12	L1ATL0, L1TRK10	L2M3	L3L3, L3TRK12	13.03 - 13.11
MUH3_LM15	L1ATN0, L1ITRK10	L2M0	L3L15	13.03 - 13.23
MUH3_LM3_TK10	L1ATN0, L1ITRK10	L2M0	L3L3, L3TRK10	13.03 - 13.11
MUH3_LM6_TK12	L1ATN0, L1ITRK10	L2M0	L3L6, L3TRK12	13.20 - 13.23
MUH3_LM10_TK12	L1ATN0, L1ITRK10	L2M0	L3L10, L3TRK12	13.03 - 13.90
MUH4_LM15	L1WTT0	L2M5	L3L15	13.03 - 13.11
MUH4_TK10	L1WTT0	L2M5	L3TRK10	13.03 - 13.11
MUH5_LM15	L1BTT0, L1TRK10	L2M5	L3L15	13.03 - 14.93
MUH6_TK10	L1WLL10, L1TRK10	-	L3TRK10	13.03 - 13.11
MUH6_LM15	L1WLL10, L1TRK10	-	L3L15	13.03 - 14.90
MUH6_TK12_TLM12	L1WLL10, L1TRK10	-	L3L0, L3CML12, L3TRK12H10	13.20 - 14.90
MUH7_TK10	L1WTL0	L2M5	L3TRK10	13.03 - 13.11
MUH7_TK12	L1WTL0	L2M5	L3TRK12(H10)	13.20 - 14.93
MUH7_LM15	L1WTL0	L2M5	L3L15	13.03 - 14.90
MUH8_ILM15	L1WTL10, L1TRK10	-	L3L15, ISO_MUON_CAL3 ISO_MUON_CM_L1	14.60 - 14.93
MUH8_TK12_TLM12	L1WTL10, L1TRK10	-	L3L0, L3CML12, L3TRK12H10	14.60 - 14.93

Table 4.1: *Definitions (L1, L2, L3) and associated periods of activity (defined in terms of trigger list versions) of all the triggers that have been studied. "-" refers to the absence of trigger condition.*

4.1.2 Choice of the triggers to be combined

A collection of 23 triggers has been kept for the OR efficiency calculation, and the complete Run IIa dataset has been split into 5 periods, as listed in Table 4.2. These periods corresponds to the main epochs of data taking. No significant change of the DØ detector occurred within these main epochs.

The triggers with names beginning with the prefixes MUH2 and MUH3 (referred to as MUH2* and MUH3*, and described in Table 4.1) have not been included in the combination because their L1 terms include a L1 central tracking condition (L1CTT) which could not be correctly tested in the official reconstructed data samples. A readout problem occurred during their period of activity, which prevented the L1CTT read-out from being reliable. The other triggers which include a L1CTT condition where kept in the combination, as it was possible for them to circumvent this problem, since their L1 muon conditions also contain the L1CTT requirement. Thus, by testing both the muon and the track conditions with the help of the L1 muon term, no information was missed.

Although the triggers MUH2* and MUH3* have not been included in the trigger combination, they have been considered for the selection of the data sample with which the efficiency of the trigger combination is calculated, as explained later in this chapter.

Epochs [-]	Single Muon Triggers
v8.00 - v10.03	MU_W_L2M5_TRK10, MU_W_L2M0_TRK3, MUW_W_L2M5_TRK10, MU_W_L2M0_TRK10
v10.30 - v11.04	MU_W_L2M3_TRK10, MUW_W_L2M3_TRK10, MUW_A_L2M3_TRK10
v12.00 - v12.37	MU_W_L2M3_TRK10, MUW_W_L2M3_TRK10
v13.03 - v13.11	MUH1_TK12, MUH1_TK10, MUH1_LM15, MUH4_LM15, MUH4_TK10, MUH5_LM15, MUH6_TK10, MUH6_LM15, MUH7_TK10, MUH7_LM15
v13.20 - v13.90	MUH1_TK12_TLM12, MUH1_TK10, MUH1_LM15, MUH5_LM15, MUH6_LM15, MUH6_TK12_TLM12, MUH7_TK12, MUH7_LM15
v14.00 - v14.51	MUH1_TK12_TLM12, MUH5_LM15, MUH6_LM15, MUH6_TK12_TLM12, MUH7_TK12, MUH7_LM15, MUH1_ILM15
v14.60 - v14.93	MUH1_TK12_TLM12, MUH1_ILM15, MUH5_LM15, MUH6_LM15, MUH6_TK12_TLM12, MUH7_TK12, MUH7_LM15, MUH8_TK12_TLM12 MUH8_ILM15

Table 4.2: *Complete list of the 23 single muon triggers that have been selected for the combination. The table is separated in five main data taking epochs (defined in terms of trigger list versions).*

4.2 Calculation of the combined trigger efficiency

4.2.1 Trigger condition tests

The tag-and-probe method

The computation of the OR relies on the test of each trigger condition included in the definitions of the combined triggers. The tag-and-probe method has been used to perform these tests, in a similar way as in Section 3.3.

We first select one offline muon of good quality (high p_T, isolated), which we refer to as the tag muon. A second offline muon is selected in the same event by requiring the invariant mass composed by the four vectors of both muons to be close to the Z boson mass. The additional muon, which is this way associated to the tag muon, is referred to as the probe muon. Each event containing a tag and a probe muon are $Z \rightarrow \mu\mu$ candidate where at least two muons should be present.

When testing a single muon trigger term, we require a probe muon to match spatially an online muon which fulfills the trigger term requirements. A matched online muon should exist since the test is performed with a $Z \rightarrow \mu\mu$ candidate event. The result of the matching is a boolean, *true* in case of success, and *false* otherwise. The test of a muon trigger term with the tag-and-probe method is represented in Figure 4.1.

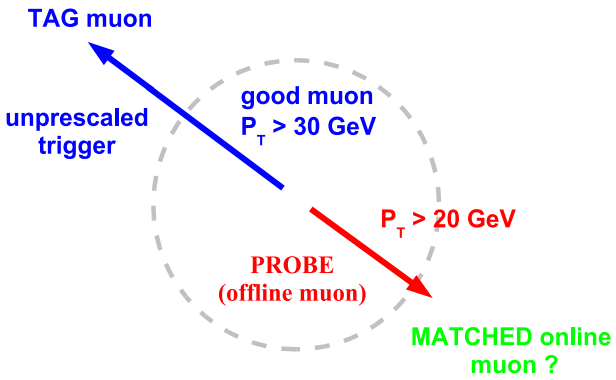

Figure 4.1: *Schematic view of the tag-and-probe method, as configured for the test of a muon trigger term. The reconstructed probe muon is required to match an online muon.*

Inclusion of prescales inefficiency

The choice has been made to include the prescales in the trigger efficiency in order to be able to determine the integrated luminosity. During Run IIa, prescales have been applied only at level 1.

The prescaler takes the decision of whether keeping an event or throwing it out. For each period of runs, the triggers and their associated prescales are defined in the "trigger list" configuration. Prescales are allowed to vary within a given trigger list period, while it is forbidden within a single run.

In the offline reconstructed data file, the information about the success or the failure of a trigger decision is stored in what we refer to as the "trigger fire bit". This bit is set to one if each trigger level condition is fulfilled and if no event rejection was operated by the prescaler. As a result, including the test of the "trigger fire bit" in the trigger efficiency calculation will account for the inefficiency caused by the prescales.

Since it accounts for the prescales, the provided OR efficiency should only be applied to data samples which exactly contain the same runs as those used for the efficiency calculation. Indeed, one cannot assume that an efficiency based on a given set of runs and their associated prescales could describe another set of runs, with potentially different prescales. As well, the statistics available for the test of a given trigger will be restricted to the runs where this trigger actually operated, and no use will be made of information coming from other triggers which include the same requirements.

Strategy to unbias the tag selection

The efficiency, as estimated by the tag-and-probe method, can be written as follows:

$$\mathcal{E}_{tag\,\&\,probe} = \frac{< \mathcal{E}_{tag\,selection} \times \mathcal{E}_{probe\,preselection} \times \mathcal{E}_{probe\,test} >}{< \mathcal{E}_{tag\,selection} \times \mathcal{E}_{probe\,preselection} >} \tag{4.1}$$

If there are correlations between the tag and the probe muons, then $\mathcal{E}_{tag\,\&\,probe}$ differs from $< \mathcal{E}_{probe\,test} >$. Such correlations can be introduced when the tag selection efficiency varies due to factors that influence the probe in the same direction. Prescales are thus perfect candidates to bias the efficiency calculation.

To unbias the tag selection with respect to the prescales, we select only tag muons that are triggered by one or several unprescaled triggers (as represented in Figure 4.1). At run level, this condition is sufficient to avoid correlations between the tag and the probe muons. However, when considering groups of runs, i.e., trigger list periods, the problem is still present if the unprescaled triggers which select the tag muon are not always the same. The chosen solution was to split the efficiency calculation into periods of runs where the unprescaled triggers which select the tag remain unchanged.

Nevertheless, the drawback of this method is the decrease of statistics for the tag muons. Indeed, in order to construct such periods, some runs need to be removed, otherwise one would be left with some very small periods where no efficiency calculation could be possible, due to high statistical uncertainties. The difficulty is thus to build "unbiased" periods and simultaneously keep enough statistics.

Epochs [–]	Unprescaled Trigger(s)	Number of runs	Number of tag muons
v8.00 - v10.03	MU_W_L2M5_TRK10	199	1673
v10.30 - v11.04	MUW_W_L2M3_TRK10	397	6811
v12.00 - v12.37	MUW_W_L2M3_TRK10	973	21532
v13.03 - v13.11	MUH2_LM3_TK12 MUH2_LM15 MUH3_LM3_TK10 MUH3_LM15	108	3105
v13.20 - v13.90	MUH1_TK12_TLM12 MUH1_LM15	949	29321
v14.00 - v14.51	MUH1_TK12_TLM12 MUH1_ILM15	408	12933
v14.60 - v14.93	MUH8_TK12_TLM12 MUH8_ILM15	459	15020

Table 4.3: *List of the chosen split of Run IIa and the associated unprescaled triggers used for the tag selection. The number of runs selected in each period and the number of tag muons available are provided.*

In the split of Run IIa proposed in Table 4.2, each period contains one or several unprescaled triggers which almost always remains unprescaled. All epochs have been analyzed in detail, run by run, in order to manage the more suitable choice of runs. The list of unprescaled triggers chosen for the tag selection[1] is given in Table 4.3. The numbers of runs and tag muons per period are also provided.

[1]MUH2* and MUH3* triggers have been kept for the tag selection of the period [v13.03, v13.11], despite the L1CTT issue. The reason lies in the fact that they constitute the best set (in term of variety of runs) of unprescaled triggers for this period. The L1CTT problem does not introduce a bias for the tag but reduces the statistics of the selection. Choosing another set of single muon triggers would have led to a larger loss of statistics.

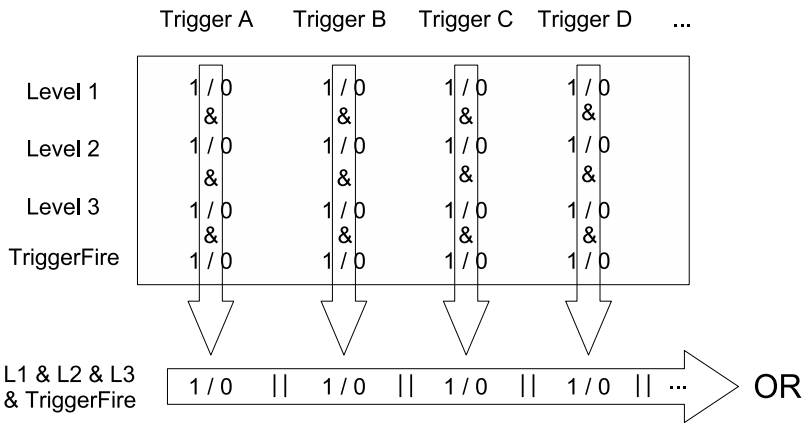

Figure 4.2: *Diagram representing the method used for the combination of the 23 single muon triggers. All the levels of a given trigger are tested separately, and the each trigger is required to have passed the prescaled (through the test of the "TriggerFire" bit). The resulting trigger tests are combined in the final stage.*

4.2.2 Combination of the triggers

Overall efficiency

A usual strategy for the calculation of a trigger efficiency consists in multiplying the conditional efficiencies relative to each (independent) level of the trigger. However, when combining many triggers, this method quickly becomes complicated. For the efficiency calculation, the choice has been made to directly combine the tag-and-probe tests of the triggers terms, without calculating its specific efficiency. As described in Figure 4.2, the test of one trigger consists in the tests of all its trigger levels, plus the test of the "trigger fire bit", which ensures that the trigger actually fired. The final combination consists in an inclusive OR of the tests of all the chosen triggers.

Choice of the parametrization

By merging the tests of different trigger levels and including the prescaler decision, we combine terms which depend on different parameters. Indeed, the tracking terms efficiencies will typically vary with the muon z position at the vertex and η_{cft} (η as measured in the CFT detector), the efficiencies of the muons terms are determined as function of $\eta_{detector}$ and ϕ, the muon isolation terms efficiencies will depend on the number of jets and the muon p_T, and the prescales are likely to be tightened at high luminosity.

Since it is not reasonable to parametrize the combined trigger efficiency with respect to more than two dimensions due to statistical limitations, we decided to only consider the $\eta_{detector}$ and ϕ parameters. Only probe muons with a p_T larger than 20 GeV have been considered for the calculation of the OR efficiency, which does not completely cancel the dependency in p_T. A systematical error has therefore been derived to account for the remaining dependence in p_T.

Depending on runs and epochs, the DØ detector could be in a different running condition. The way Run IIa has been split should prevent the efficiencies from such biases. Furthermore, the OR efficiencies also depend on runs because of the prescales they include. As long as the analyzer

only considers data taken during the same runs as those used for the OR efficiency calculation, averaging on the run numbers will have no consequence.

The dependence on z can be neglected as long as the OR efficiency is used in an analysis where the data shows a vertex profile in z similar to the $Z \to \mu\mu$ sample on which the OR efficiency is based, which is the case for the leptoquark search described in this thesis.

4.3 Uncertainties of the calculation

4.3.1 Statistical uncertainties

The relative symmetrical statistical errors are available for each single bin of each epoch. Within a given epoch, one bin is independent of another bin. As well, one bin in a given epoch is independent of the same bin in another epoch. As a result, given that one selects exactly one muon per event in an analysis, the relative statistical error on the OR efficiency, can be written as follows:

$$\Delta\mathcal{E} = \sqrt{\sum_{(\eta,\phi)} \sum_{epoch} \left(\frac{\mathcal{L}_{epoch}}{\mathcal{L}} \frac{\mathcal{N}_{(\eta,\phi)}}{\mathcal{N}} \Delta\mathcal{E}_{(\eta,\phi),epoch} \right)^2} \tag{4.2}$$

where $\mathcal{N}_{(\eta,\phi)}$ is the number of selected events containing a muon with the given $\eta_{detector}$ and ϕ, $\mathcal{N}$ is the sum of all the selected events, $\mathcal{L}_{epoch}$ is the integrated luminosity of one epoch, $\mathcal{L}$ is the total integrated luminosity, and $\Delta\mathcal{E}_{(\eta,\phi),epoch}$ is the relative statistical error of one bin, for one given epoch.

The individual $\Delta\mathcal{E}_{(\eta,\phi),epoch}$ errors can be relatively large depending on the epoch. Nevertheless, we found a relative statistical error of 0.3% for the leptoquark candidate events selected in Chapter 5, as expected by the quadratic sum of many terms.

4.3.2 Systematical uncertainties

Level 3 muon central matching term

In trigger list v13, the information about the level 3 muon central matching (L3CM) term is not available in the official data set. Thus, no matching test of this term has been done explicitly in the OR efficiency calculation, for this epoch. Nevertheless, we require for each trigger the "trigger fire bit" to be 1, which includes a test of this missing term but does not ensure that the trigger which fired is matched to the probe muon.

We calculated a systematical error for this missing term with the help of a reprocessed data sample in which the information about this term was correctly filled. By comparing the OR efficiency in trigger list v13 with and without performing the L3CM term test, we derived a relative error lower than 0.1%, which is thus negligible.

Level 3 muon isolation term

As for the L3CM term in v13, the level 3 isolation terms are not available for the period [v14, v14.51] in the official data set. With the same method as for the L3CM term error, we derived a relative error of 0.4% for the missing level 3 isolation term.

Choice of the binning

To estimate the error relative to the choice of the binning, we first applied the provided OR efficiency to simulated events, and a second time with the same efficiency binned with two times more bins. We finally compared the number of remaining events after all cuts and derived a relative error of 0.2%.

Tag-and-probe bias

The proposed method of efficiency calculation removes the correlations between the tag and the probe muons when they are due to the prescales. However, the existence of a bias in the tag-and-probe method not due to the prescales was reported by the muon identification group [63], where an error of 0.2% has been quoted. Since we used the same tag-and-probe method as the one used for the muon identification certification, this error should apply here.

Missing muon p_T parameter

To account for the fact that the dependence in muon p_T is missing, we shifted the p_T cut of the probe muon from 20 GeV to 30 GeV and compared the resulting OR efficiencies. A relative error of 0.6% has been found.

Background contamination

The quality of the $Z \to \mu\mu$ selection used of the OR efficiency calculation is sensitive to the background contamination due to QCD multijet events or $W \to \mu\nu$ events. To further reduce this background, we tightened some cuts of the $Z \to \mu\mu$ events selection and recalculated the OR efficiency (with the loosest offline muon and track qualities).

- By constraining the tag and the probe muons to be back to back ($\Delta\Phi > 2.9$), we found a variation of 0.3%.
- Tightening the *dca* cut from $|dca| < 0.2$ cm to $|dca| < 0.1$ cm led to a variation of 0.2%.
- When requiring to have no reconstructed jet ($p_T > 15$ GeV), a variation of 0.2% was observed.

4.4 Properties of the combined triggers efficiency

4.4.1 Improvement compared to single muon triggers efficiencies

The gain of the OR efficiency compared to single muon triggers efficiencies is clearly highlighted in Figure 4.3. In addition to the overall improvement of the efficiency in the central and wide η

regions, the acceptance is increased in the forward region ($1.0 \leq \eta < 2.0$), thanks to the trigger MUH5_LM15.

The Figure 4.3 also represents the efficiencies of MUH1_TK12_TLM12 and MUH1_LM15, which have been chosen as unprescaled triggers for the tag selection of the period [v13.2, v13.9]. Since these triggers are unprescaled, and because of the complementariness of their L3 conditions (muon and tracking criteria), their contribution to the OR is dominant, as emphasized by the figure.

4.4.2 Dependencies

As shown on the left plot of Figure 4.4, the OR efficiency depends on the offline quality of the probe muon with respect to which it is calculated, especially in the central η region. For tighter muon qualities, the efficiency increases and its η dependence becomes flatter. The muon qualities are described in Section 3.2.2.

At low luminosities, triggers are less sensitive to prescales, and thus offer their larger contribution to the OR efficiency. For higher luminosities, the efficiencies of the OR and of the single muon triggers are decreasing significantly, as they are more likely to be prescaled. The right plot of Figure 4.4 represents these variations. It also shows efficiencies which do not include prescaler decisions (the dashed plots), and which therefore do not decrease with the luminosity. Another representation of the inefficiency associated with the prescales is provided in Figure 4.5, where OR efficiency maps with and without prescaler decisions are compared.

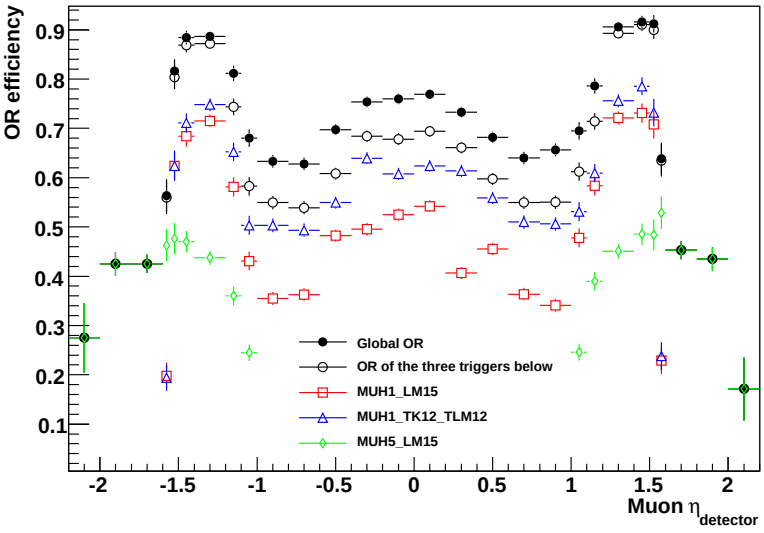

Figure 4.3: *Dependence of the single muon trigger OR efficiency as function of $\eta_{detector}$, for the data taking period [v13.2, v13.9]. The efficiencies of some combined triggers are also represented. All efficiencies are given with respect to a loose muon matched to a medium track, and fulfilling the loose isolation requirement. The muon qualities are described in Section 3.2.2.*

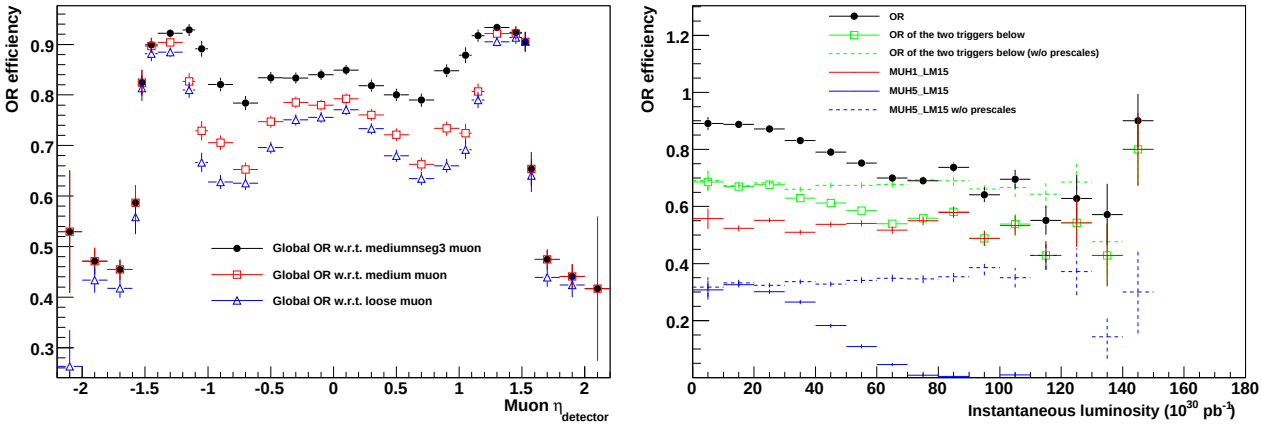

Figure 4.4: *Left plot: dependence of the single muon OR efficiency as function of $\eta_{detector}$, and with respect to three offline muon qualities: loose, medium, and mediumnseg3 (medium muon found with hits in the 3 layers of the muon detector). Right plot: dependence of the single muon OR efficiency as function of the instantaneous luminosity. For both plots, efficiencies are calculated with data collected during the trigger list period [v13.2, v13.9], and given with respect to a mediumnseg3 probe muon matched to a medium track, and fulfilling the tight isolation requirement. The muon qualities are described in Section 3.2.2.*

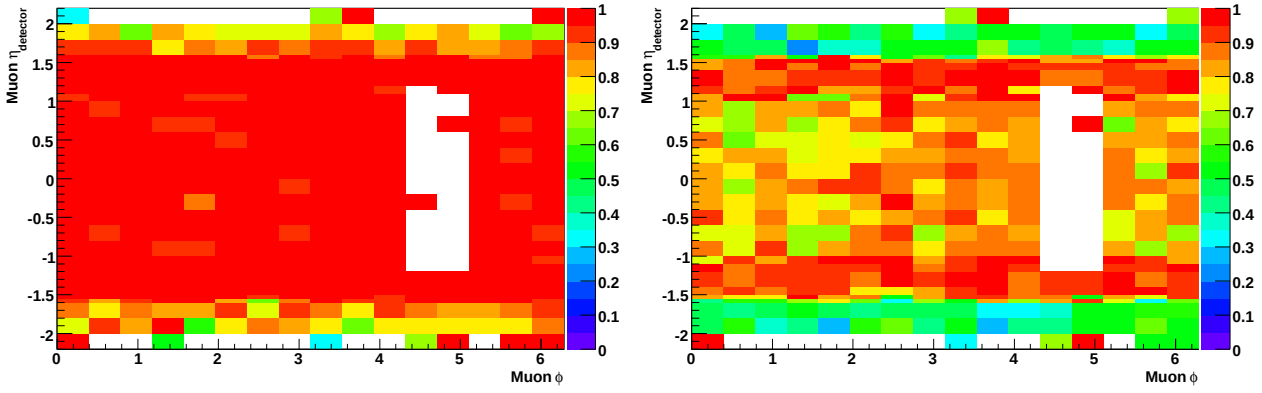

Figure 4.5: *(ϕ,$\eta_{detector}$) map of the OR efficiency. In the left plot, the prescales are not taken into account, whereas in the right plot, they are included in the efficiency. For both plots, efficiencies are calculated with data collected during the trigger list period [v13.2, v13.9], and given with respect to a mediumnseg3 probe muon matched to a medium track, and fulfilling the tight isolation requirement. The muon qualities are described in Section 3.2.2.*

4.4.3 Averaged efficiency

Figure 4.6 shows the OR efficiency as function of the trigger list versions, and averaged over all other parameters.

The OR efficiency is higher in the period [v13.03,v13.90], which corresponds to the period where the largest number of triggers has been combined. In addition, this epoch is the only one to include a trigger which covers the forward region.

In the periods prior to trigger list v12, only one trigger is used, which explains the lower efficiency. Furthermore, during these epochs, only limited luminosity was available, and thus, the efficiency suffers from statistical fluctuations.

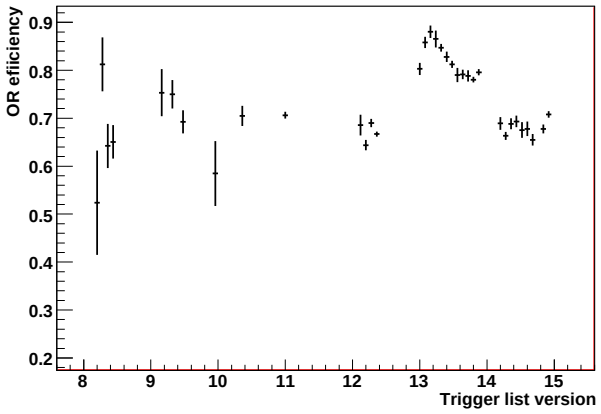

Figure 4.6: *OR efficiency as function of the trigger list versions, and averaged over all other parameters.*

4.5 Data selection and luminosity calculation

4.5.1 Removal of runs not used in the efficiency calculation

As explained previously, the fact of including the prescales in the OR efficiency implies that one cannot a priori extrapolate the efficiency obtained from a given set of runs to a larger set of runs. Due to the method used for the tag selection, not the complete Run IIa but an important subgroup of its runs have been selected for the efficiency calculation (nearly 95% of the complete available dataset). To be consistent, only the runs used for the efficiency calculation are considered for the leptoquark search analysis. One additional run has been removed from this list of runs because of how the luminosity is calculated, as described further in this section.

4.5.2 Online-to-offline muons matching and trigger selection

The OR efficiency has been estimated with the tag-and-probe method, which relies on the matching between an offline muon and an online one. As a consequence, in order to be able to describe data events by applying this efficiency to simulated events, one should ensure that the selected data events contain an offline muon (of the desired quality) which is matched to at least one of the

combined triggers. One should also check that the trigger for which the online-to-offline matching conditions are fulfilled actually fired, since it is also required in the efficiency calculation.

In the case of the $\mu \not{E}_T jj$ channel, an overall matching inefficiency of 4.1% has been observed after the application of the preselection cuts (as defined in Section 5.1.1). Figure 4.7 shows the online-to-offline matching efficiencies of the five data taking periods considered.

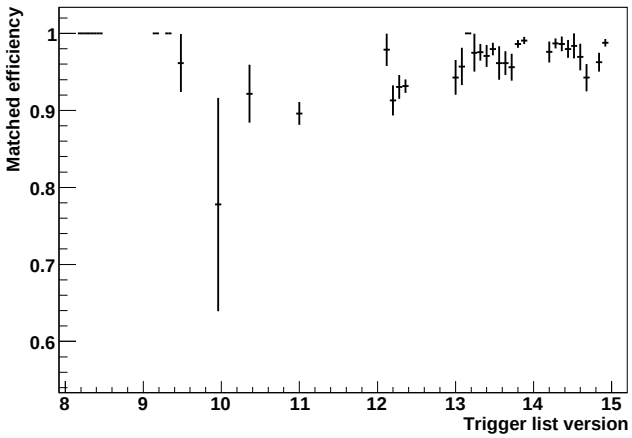

Figure 4.7: *Online-to-offline matching efficiencies in the $\mu \not{E}_T jj$ channel as function of the trigger list versions. Results are shown after the application of the preselection cuts.*

4.5.3 Luminosity calculation

The calculation of the OR efficiency has been complicated by the inclusion of the prescales because of the biases they introduce in the tag-and-probe method. Nevertheless, the advantage of this method lies in the luminosity calculation, which is made possible even though triggers with several prescales are combined.

In the list of runs considered for the OR efficiency calculation, which has to be used in an analysis, each run includes at least one unprescaled single muon trigger. This implies that the integrated luminosity could be calculated with any unprescaled trigger, and even only one, given that it is configured for all the runs of the selected data sample. The choice has been made to estimate the luminosity with the trigger JT_125TT, since it is almost always unprescaled. This trigger is designed to fire on jet events, given that its transverse momentum is greater than 125 GeV. Since the probability that such an event occurs is rather low, this trigger does not need any prescale.

After removing the bad runs defined by the official data quality group, only 6 runs where JT_125TT is prescaled remained, and only one of these 6 (the run number 208850) was included in the OR efficiency computation. This run has been removed from the list of runs used for the OR efficiency calculation, and the resulting list has to be considered as a "good runs list" for the estimation of the luminosity.

More than 90% of the luminosity delivered by the Tevatron is actually recorded by DØ, which is represented in Figure 2.9. Using the standard run quality conditions, an integrated luminosity of $1061.36\,\mathrm{pb}^{-1}$ is measured. In the case of the $\mu \not{E}_T jj$ channel, the results of the luminosity calculation, based on the "good runs list", are provided in Table 4.4. An integrated luminosity of

$1002.24\,\mathrm{pb}^{-1}$ is calculated, which constitutes a large fraction of the recorded luminosity. For the $\mu\mu jj$ channel, the same "good runs list" is utilized, and an integrated luminosity of $1011.60\,\mathrm{pb}^{-1}$ is determined. The variation of the luminosity between the two channels lies in the fact that differents data samples are utilized for each channel.

Epochs [–]	Unprescaled Trigger	$\mathcal{L}_{delivered}$ (pb^{-1})	$\mathcal{L}_{recorded}$ (pb^{-1})	$\mathcal{L}_{good}$ (pb^{-1})
v8.00 - v10.03	JT_125TT	100.23	84.12	24.19
v10.30 - v11.00		89.52	83.49	70.93
v12.00 - v12.37		269.32	248.11	220.03
v13.03 - v13.20		55.51	49.94	31.68
v13.20 - v13.90		404.03	372.42	324.33
v14.00 - v14.51		178.23	166.72	140.50
v14.60 - v14.93		236.91	220.60	190.59
Total		1333.75	1225.41	1002.24

Table 4.4: *Integrated luminosities corresponding to each epoch considered for the combination of triggers, in the case of the $\mu \not{E}_T jj$ channel.*

4.6 Application of the trigger efficiency to simulated events

In order to account for the trigger efficiency, a trigger weight is applied to each simulated event, depending on the $\eta_{detector}$ and ϕ values of the offline muon(s).

Since there is no trigger list version information in Monte-Carlo events, the event trigger weight is calculated as the average of the single muon trigger OR efficiency over all the periods of data taking.

4.6.1 Case of the $\mu \not{E}_T jj$ final state search

As described in Section 5.1.1, a muon veto is required, so that only events with exactly one reconstructed muon are selected. Thus, the event trigger weight can be written as follows:

$$\mathcal{W}_{trig} = \frac{\sum_{epoch} \mathcal{L}_{epoch} \times \mathcal{E}_{\mu_{(\eta,\phi)}}}{L} \tag{4.3}$$

where $\mathcal{E}_{\mu_{(\eta,\phi)}}$ is the OR trigger efficiency of a muon with the given $\eta_{detector}$ and ϕ, $\mathcal{L}_{epoch}$ is the integrated luminosity of one epoch, and L is the total integrated luminosity.

4.6.2 Case of the $\mu\mu jj$ final state search

For this channel, each event contains at least two muons, as described in Section 5.2.1. The method to calculate the event trigger weight would consist in estimating the probability $\mathcal{P}$, that none of the muons were triggered. Then $1 - \mathcal{P}$ would give the probability that at least one of the muons was triggered. In the case where triggers efficiencies account for the prescales, $\mathcal{P}$ is a priori not equal to the product of the trigger efficiencies relative to each muon of the event, and therefore cannot be calculated easily. This is due to the fact that the trigger efficiencies relative to each muon are correlated due to the prescales.

As a result, another strategy has been chosen for the trigger selection of the $\mu\mu jj$ channel. The same method as explained in this chapter has been used, but only eight unprescaled triggers have been combined to avoid possible bias due to correlations, and so that the event trigger weight could be defined by:

$$\mathcal{W}_{trig} = 1 - \frac{\sum_{epoch} \mathcal{L}_{epoch} \times \prod_{\mu(\eta,\phi)} (1 - \mathcal{E}_{\mu_{(\eta,\phi)}})}{L} \tag{4.4}$$

where the product $\prod_{\mu(\eta,\phi)}$ includes all the muon efficiencies of the event.

The eight unprescaled triggers that have been combined are listed in Table 4.5, and consist of the triggers used for the tag muon selection, excepted for the period [v13.03,v13.11], because of the L1CTT problem explained in Section 4.1.2. To cover the epoch [v13.03,v13.11], we used the trigger MUH1_TK10.

Epochs [-]	Single Muon Triggers
v8.00 - v10.03	MU_W_L2M5_TRK10
v10.30 - v11.04	MUW_W_L2M3_TRK10
v12.00 - v12.37	MUW_W_L2M3_TRK10
v13.03 - v13.11	MUH1_TK10
v13.20 - v13.90	MUH1_TK12_TLM12 MUH1_LM15
v14.00 - v14.51	MUH1_TK12_TLM12 MUH1_ILM15
v14.60 - v14.93	MUH8_TK12_TLM12 MUH8_ILM15

Table 4.5: *Complete list of the eight single muon triggers that have been selected for the combination. The table is separated in five main data taking epochs (defined in terms of trigger list versions).*

In order to check whether the use of the efficiencies calculated with the eight unprescaled triggers of Table 4.5 introduces a bias when applied to a multimuon event, we have calculated the $Z \to \mu\mu$ production cross section. Two jet inclusive $Z \to \mu\mu$ samples simulated by the ALPGEN and PYTHIA event generators (as described in Section 3.1.2) with di-muon masses between 60 and 130 GeV have been considered. No correction due to the fixed di-muon mass window has been applied. We corrected the simulated events with the reconstruction efficiencies and rescaled their contribution in order to reach agreement with data. From this, we derived the $Z \to \mu\mu$ production cross section.

With both event generators, a cross section of about 246 pb has been derived, which is to be compared with the 256.6 pb expected (calculated at NNLO QCD) [55]. The observed variation of about 4% stays within the error on the luminosity, namely 6.1%.

4.6.3 Event trigger weights

Table 4.6 provides the event trigger weights $\mathcal{W}_{trig}$ as calculated for both channels, including the efficiencies for the main background and for the signal. The weights of the $\mu \not{E}_T jj$ channel are lower than those of the $\mu\mu jj$, even if more triggers are combined, because only one muon is selected and can potentially fire a trigger. All the values are given at preselection level, as defined for both channels in Section 5.1.1 and 5.2.1.

	Simulated Sample			
Channel	$W(+jets) \to l\nu + jets$	$Z/\gamma^*(+jets) \to \mu\mu + jets$	$t\bar{t}$ inclusive	signal
$\mu \not{E}_T jj$	73.74	73.54	76.68	76.63
$\mu\mu jj$	82.65	80.77	82.98	83.11

Table 4.6: *Event trigger weights (in %) for each simulated sample type and each channel. Values are given at preselection level, as defined for both channels in Section 5.1.1 and 5.2.1. The assumed leptoquark mass is 260 GeV.*

Chapter 5

Selection of candidate events in $\mu \not{E}_T jj$ and $\mu\mu jj$ final states

This chapter presents the selection of leptoquark candidate events in the case of two pair production channels, namely the one leading to the final state composed of one muon, missing transverse energy, and two jets, and the one leading to the final state composed of two muons and two jets. Figure 5.1 shows the typical leptoquark pair production process for each considered channel. As defined in Section 1.4, β, the branching fraction for a leptoquark to decay into a muon and a quark ($\beta = Br(LQ \to \mu q)$), governs the sensitivity of the two channels.

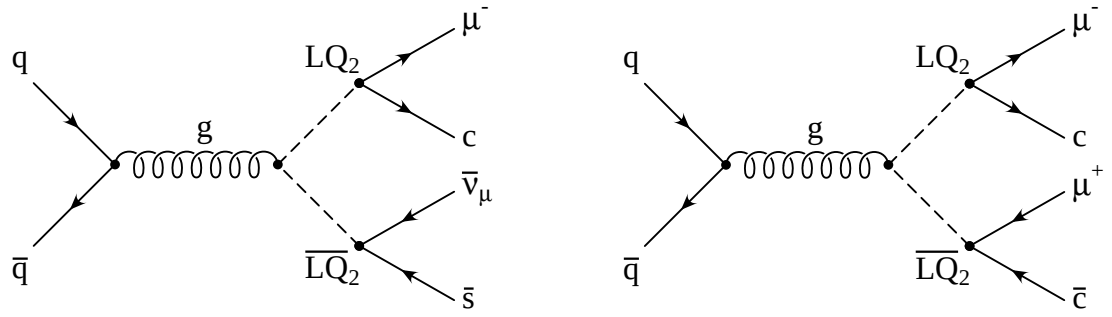

Figure 5.1: *Typical second generation scalar leptoquark pair production process leading to a final state composed of a muon, missing transverse energy, and two jets (left), or to a final state composed of two muons and two jets (right).*

The selections have been performed for leptoquark signals generated for the two different topologies, with assumed masses running from 140 to 320 GeV, and for values of β from 0.1 to 1. For $\beta = 0$ both of the analyzed channels vanish, while the channel where both leptoquarks would decay to a neutrino and a jet has a good sensitivity. Muons arising from the decay of leptoquarks in the $\mu\mu jj$ final state have a non negligible probability not to be reconstructed. $\mu\mu jj$ signal events turn out to survive the $\mu \not{E}_T jj$ channel selection cuts with up to nearly half of the efficiency for $\mu\nu qq$ signal events. By contrast, the rate that a particle included in events of the $\mu \not{E}_T jj$ topology fakes a muon can be neglected. A muon fake rate of the order of 0.03% has been estimated when requiring the event selection of the $\mu\mu jj$ channel described in Section 5.2.1. The contribution of $\mu\mu qq$ signal events to the $\mu \not{E}_T jj$ analysis has been included in a consistent

way, such that the selections of both channels remain orthogonal.

Cut-based selections have been considered for both channels in a first approach. Optimizations using a multivariate discriminant (the k-Nearest-Neighbor algorithm) have then been performed to increase the signal enhancement and the background rejection.

5.1 Selection in the $\mu \not{E}_T jj$ final state

5.1.1 Preselection

Preselection cuts

Events are required to have at least one *mediumnseg3* muon found within $|\eta|$ <2 and with a tight isolation. The selected muon is in addition required to be matched to a central track of at least medium quality and of transverse momentum exceeding 20 GeV. The definition of the object qualities are provided in Section 3.2.2. A veto on muons is placed so that not more than one loose muon fulfilling the muon preselection criteria of the $\mu\mu jj$ channel (defined in Section 5.2.1) is present in the event. Events of both channels are thus orthogonally selected, and the preselected $\mu \not{E}_T jj$ events only contain one *mediumnseg3* muon with the constraints described above. At least two jets are required to deposit energy in the region $\eta < 2.5$ of the calorimeter, with transverse energies greater than 25 GeV. Events with a missing transverse energy larger than 30 GeV are selected, which reduces the contribution of QCD multijet events.

The azimuthal angle ($\Delta\phi$) between the muon and the missing transverse energy is required to be lower than 3 radians, in order to remove events with badly reconstructed muons. Such events contain an overestimated $\not{E}_T$, since the $\not{E}_T$ is corrected for muons. (see Section 3.2.3) This lies in the fact that an overestimated muon p_T contributes significantly to the direction of the $\not{E}_T$, namely in the opposite direction with respect to the muon.

The transverse mass $M_T(\mu, \not{E}_T)$ reconstructed from the four-vector of the muon and the missing transverse energy is required to exceed 50 GeV. The transverse mass of two objects separated in the azimuthal plane by the angle ($\Delta\phi$), and of transverse momenta respectively p_{T_1} and p_{T_2}, is given by the following formula:

$$M_T = \sqrt{2 \times p_{T_1} \times p_{T_2} \times (1 - cos(\Delta\phi))} \tag{5.1}$$

The motivation for this cut is to remove a large part of the QCD multijet background. In addition, events with $M_T(\mu, \not{E}_T)$ lower than 50 GeV are not in the kinematic region of the presented search.

All the reconstruction cuts plus the precedent cuts are referred to as the preselection cuts. After these cuts, 5744 background events, and 5693 data events are remaining, as listed in Table 5.2. We define the signal efficiency as the ratio of the remaining number of weighted events over the number of generated events. In this definition, events are weighted as described in Section 3.4, but do not include the channel branching ratio. 23.5% and 10.6% of $\mu\nu qq$ and $\mu\mu qq$ signal events, respectively, are surviving the preselection cuts (see Table 5.2).

Estimation of QCD multijet background and background rescaling

The background of QCD multijet events with a non-prompt muon has been estimated from the data sample after all preselection cuts, except those on $\not{E}_T$ and $M_T(\mu, \not{E}_T)$, and by requiring a

reverted tight isolation cut on the muons instead of an isolation. The reverted tight isolation cut is defined by turning the upper bounds on the calorimeter and track isolation (defined in Section 3.2.2) into lower bounds. It thus requires the sum of transverse momenta of all tracks in a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.5$ around the muon to be larger than 2.5 GeV, and the energy deposited in the calorimeter in an annulus of radius $0.1 < \Delta R < 0.4$ around the muon to exceed 2.5 GeV. The resulting QCD sample is used to describe the kinematics of multijets events.

In order to determine the rate of the QCD multijet contribution, we further normalize the QCD sample to data in the region where QCD multijet events are dominant, i.e. for $\not{E}_T < 10$ GeV. In this region of $\not{E}_T$, semi-leptonic b quark decays are contributing to the $\mu \not{E}_T jj$ final state. The normalization factor F_{QCD} is obtained after removing the contribution of the other backgrounds (W/Z bosons and $t\bar{t}$ productions).

In parallel, the W, $Z/\gamma^* \to \mu\mu$ and $Z/\gamma^* \to \tau\tau$ samples are simultaneously normalized to data by the factor $F_{W/Z}$, which is derived at preselection level, in the region of the W boson peak ($M_T(\mu, \not{E}_T) \in]50,\ 110[$ GeV). The choice of normalizing the contribution of these backgrounds by the same factor is motivated by their similar production mechanism. In addition, the contribution of the $t\bar{t}$ production in the resonance region of the W boson contribution is negligible compared to the W sample rate.

To rescale both the QCD and the "W/Z" samples, we solve the following linear system:

$$\begin{pmatrix} N_{data} - N_{t\bar{t}} \\ M_{data} - M_{t\bar{t}} \end{pmatrix} = \begin{pmatrix} N_W + N_{Z\mu\mu} + N_{Z\tau\tau} & N_{QCD} \\ M_W + M_{Z\mu\mu} + M_{Z\tau\tau} & M_{QCD} \end{pmatrix} \times \begin{pmatrix} F_{W/Z} \\ F_{QCD} \end{pmatrix}$$

where:

$$\begin{cases} N_{\text{i}} \text{ is the number of events for } \not{E}_T < 10 \text{ GeV}, \\ M_{\text{i}} \text{ is the number of events in the bulk of the } W \ (\ M_T(\mu, \not{E}_T) \in]50,\ 110[\text{ GeV}) \end{cases}$$

We find $F_{QCD} = 0.0099$ and $F_{W/Z} = 1.50$, given that all Monte-Carlo backgrounds are initially normalized to data by using NLO or NNLO cross section predictions, as described in Section 3.1.2. Figure 5.2 shows the distributions of the missing transverse energy and the $M_T(\mu, \not{E}_T)$ mass in their normalization windows, as well as the missing transverse energy distribution with a wide range, and the $M_T(\mu, \not{E}_T)$ transverse mass given the condition $\not{E}_T < 10$ GeV. The correct Data/MC agreement of $M_T(\mu, \not{E}_T)$ and $\not{E}_T$ outside their respective normalization windows demonstrate the validity of the QCD estimation method, which theoretically suffers from the possible variations of the QCD shape.

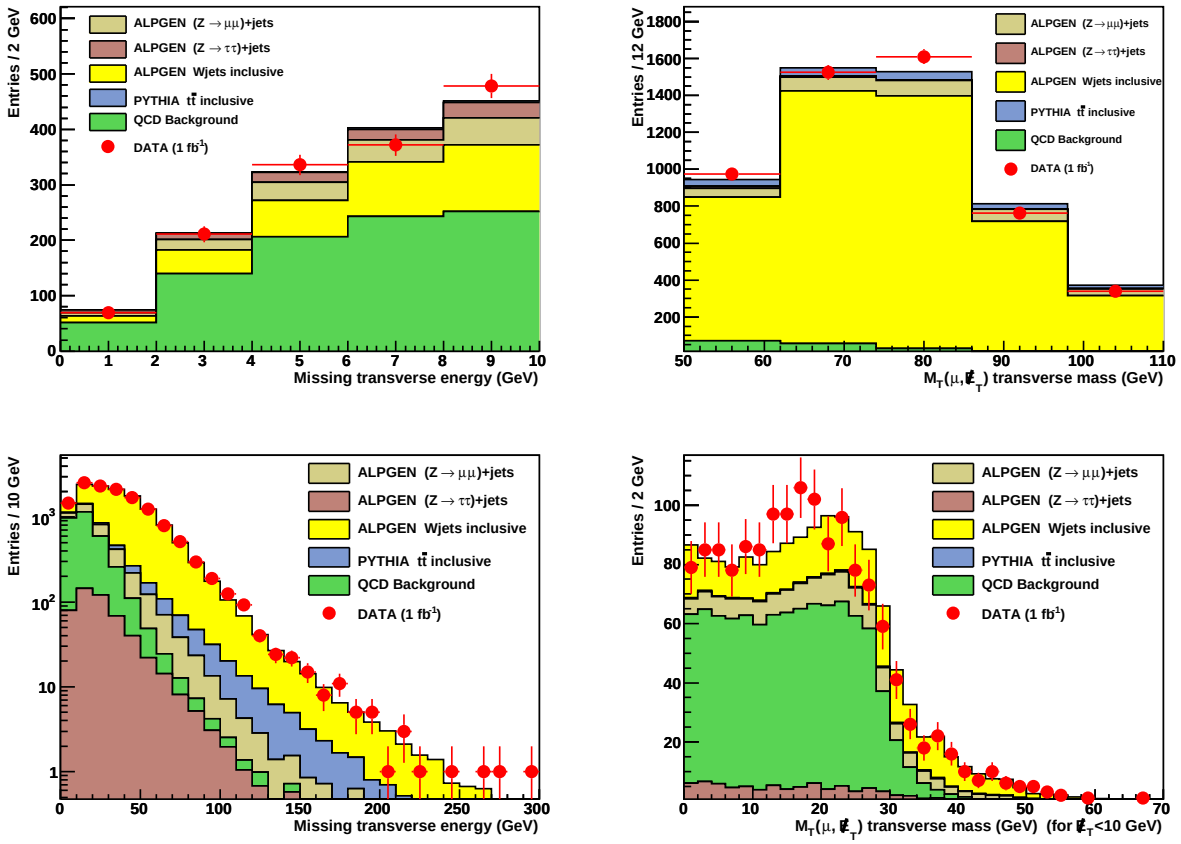

Figure 5.2: *Estimation of the QCD multijet background. The upper left and right plots show the $\not\!\!E_T$ and $M_T(\mu, \not\!\!E_T)$ distributions, respectively, in their normalization windows. The lower left plots represent the measured $\not\!\!E_T$ over a wide range, and the lower right plot provides the transverse mass $M_T(\mu, \not\!\!E_T)$ given the condition $\not\!\!E_T < 10\,GeV$. All distributions are shown after all preselection cuts, except those on $\not\!\!E_T$ and $M_T(\mu, \not\!\!E_T)$. The QCD background is selected with a reverted muon isolation cut. The data is represented by red dots, and the Standard Model expectation by colored histograms.*

Comparison of the Data with the Standard Model expectation

The plots of Figure 5.3 compare the data with the Standard Model expectation at preselection level, for several distributions. The muon and the two leading jets transverse momenta, as well as the muon η, are well described by the Standard Model background. The transverse mass $M_T(\mu, \not{E}_T)$, the missing transverse energy (represented on Figure 5.2) and its separation in ϕ with respect to the muon, which are variables used for the preselection, also reach agreement with the theoretical prediction.

5.1.2 Cut-based analysis

The following cut-based analysis focuses on the hypothesis that the branching fraction β is equal to 0.5. As explained in Section 1.4, this value maximizes the signal rate of $\mu\nu qq$ generated events.

Discriminating variables and cut optimization

In order to enhance the signal contribution and reject the background from Standard Model processes, we consider three selection variables:

- The transverse mass $M_T(\mu, \not{E}_T)$, in order to suppress the resonance of the W boson.
- The scalar transverse variable S_T, defined as the sum of all the final state particle transverse momentum:
$$S_T = P_T^{\mu_1} + P_T^{jet_1} + P_T^{jet_2} + \not{E}_T \tag{5.2}$$
This variable is expected to be larger for the signal as the leptoquark decay products are more energetic than those of the backgrounds.
- The transverse mass $M_T(\not{E}_T, jet_1)$, constructed from the four-vector of the leading jet (with respect to the transverse momentum p_T) and the missing transverse energy. This quantity has been chosen due to its relation to the reconstructed leptoquark mass, and the good separation it provides between signal and background. The variables which contribute to the reconstruction of the leptoquark mass are the transverse or invariant masses which are constructed from the four-vectors of the muon and one of the two leading jets, and from the $\not{E}_T$ and the four-vector of one of the two leading jets.

Figure 5.4 represents the $M_T(\not{E}_T, jet_1)$ and S_T distributions at preselection level. The transverse mass $M_T(\mu, \not{E}_T)$ is provided in Figure 5.3, also with the preselection cuts applied. The three plots show good agreement between the data and the Standard Model expectation, as well as a good separation between the signal and the background.

Since the preliminary version of this analysis [65] showed no evidence for leptoquarks, we optimized the cuts on the selection variables by minimizing the upper bound on the expected leptoquark pair production cross section. The expected cross section upper bound describes how well the experiment can exclude signal given that only background is observed. The best exclusion is obtained when the maximal value of the cross section is as low as possible. The limit calculation has been performed following the description in Section 6.1.

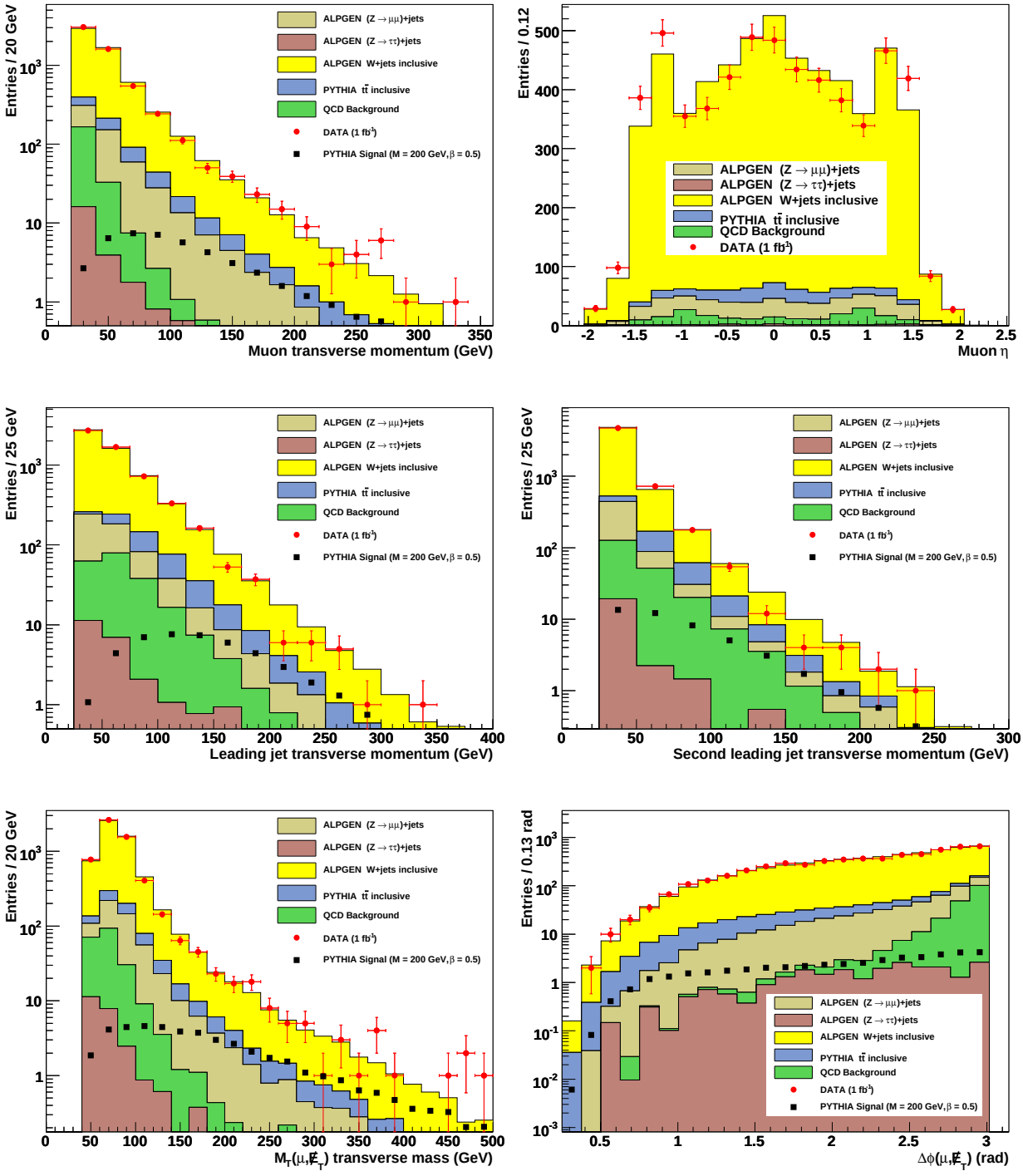

Figure 5.3: *Comparison between the data (red dots) and the Standard Model expectation (colored histograms) for several distributions: the p_T of the leading jet, the second leading jet and the muon, the muon η, the azimuthal opening angle between the muon and the $\not{E}_T$, and the transverse mass $M_T(\mu, \not{E}_T)$. The black squares describe the leptoquark signal, generated assuming $M_{LQ,gen} = 200\ GeV$ and $\beta = Br(LQ \to \mu q) = 0.5$.*

Expected upper limits at 95% confidence level on the leptoquark pair production cross section have been calculated for each ($M_T(\mu, \not{E}_T)$, S_T, $M_T(\not{E}_T, jet_1)$) point in the space limited by:

$$\begin{cases} \frac{M_T(\mu, \not{E}_T)}{\text{GeV}} & \in [130, 170] \\ \frac{S_T}{\text{GeV}} & \in [350, 500] \\ \frac{M_T(\not{E}_T, jet_1)}{\text{GeV}} & \in [150, 200] \end{cases}$$

The optimized point corresponds to the minimal expected cross section limit, which is obtained for $M_T(\mu, \not{E}_T) > 130\,\text{GeV}$, $S_T > 390\,\text{GeV}$, and $M_T(\not{E}_T, jet_1) > 180\,\text{GeV}$. Only the leptoquark signal generated with a mass of 200 GeV has been considered for this optimization, which has a minimal impact on the sensitivity of the selection, since the obtained minimal cross section limit (0.18 pb), when interpreted as a lower bound on the leptoquark mass, corresponds to a lower limit of about 200 GeV.

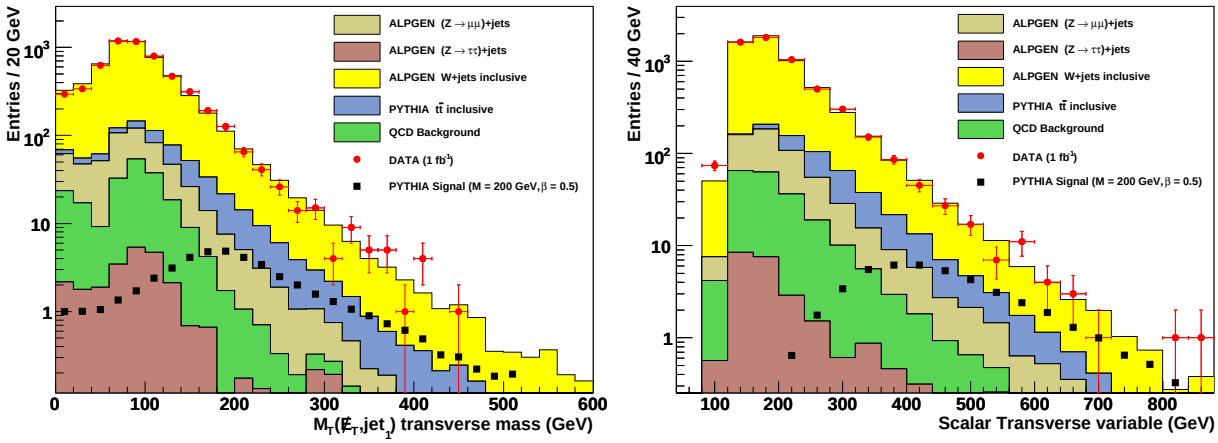

Figure 5.4: *Distributions of $M_T(\not{E}_T, \mu)$ and S_T at preselection level. Data (red dots), Standard Model background (colored histograms), and signal events (black squares) are represented. The leptoquark signal is generated assuming $M_{LQ,gen} = 200$ GeV, and shown in the case $\beta = Br(LQ \to \mu q) = 0.5$.*

Selected events and signal efficiencies

Tables 5.1 and 5.2 provide the signal efficiencies after all cuts for the various assumed leptoquark masses, and the cut flow including the preselection and each selection cut, respectively. After all the cuts, events arising from W boson and $t\bar{t}$ production turn out to be the main backgrounds to the leptoquark signal. 8.6 data events are surviving the cuts, while 8 expected background events are remaining. 7.28% and 10.6% of $\mu\nu qq$ and $\mu\mu qq$ signal events are surviving the preselection cuts, respectively.

Since no excess of data over background has been observed, upper limits on the leptoquark pair production cross section have been derived for each assumed leptoquark mass. An interpretation of the results is given in Section 6.2.1. The systematic uncertainties on the number of

remaining events are taken into account in the limit calculation. In the case $M_{LQ} = 200\,\text{GeV}$ and $\beta = 0.5$, they are listed in Table 5.2. The description of the systematic errors of the cut-based analysis is provided in Section 5.1.4, together with the uncertainties of the multivariate analysis, which we introduce in the following section.

M_{LQ} (GeV)	$\epsilon_{Signal}^{\mu\nu qq}$ (%)	$\epsilon_{Signal}^{\mu\mu qq}$ (%)
140	1.83	0.842
160	3.23	1.31
180	4.99	2.03
200	7.28	2.86
220	9.58	3.89
240	11.6	4.73
260	13.1	5.35
280	14.5	5.97
300	15.6	6.64
320	16.5	6.86

Table 5.1: *Signal efficiencies after all cuts, for leptoquark decays into $\mu\nu qq$ and $\mu\mu qq$ final states, and for all the considered assumed leptoquark masses. The branching fraction β is taken equal to 0.5.*

Samples	Preselection	$M_T(\mu, \not{E}_T)$ >130 GeV	$S_T > 390$ GeV	$M_T(\not{E}_T, jet_1) > 180$ GeV
$W(+jets) \to l\nu + jets$	4953 ± 23 ± 351	215 ± 4 ± 30	36 ± 1 ± 6	4.7 ± 0.5 ± 1.0
$Z/\gamma^*(+jets) \to \mu\mu + jets$	370 ± 3 ± 26	32 ± 1 ± 4	5.6 ± 0.3 ± 0.8	0.83 ± 0.09 ± 0.13
$Z/\gamma^*(+jets) \to \tau\tau + jets$	24 ± 2 ± 2	0.88 ± 0.26 ± 0.12	0.31 ± 0.17 ± 0.05	0.014 ± 0.003 ± 0.002
$t\bar{t}$ (inclusive)	214 ± 1 ± 39	23.8 ± 0.3 ± 4.6	7.4 ± 0.2 ± 1.5	2.5 ± 0.1 ± 0.5
QCD multijet	187 ± 1 ± 37	3.9 ± 0.2 ± 0.8	1.0 ± 0.1 ± 0.2	0.49 ± 0.07 ± 0.10
Total Background	5748 ± 23 ± 394	276 ± 4 ± 37	50 ± 1 ± 8	8.6 ± 0.6 ± 1.4
Data	5693	263	52	8
$\epsilon^{\mu\nu qq}_{Signal}$ (%)	23.5	14.8	10.8	7.28
$\epsilon^{\mu\mu qq}_{Signal}$ (%)	10.6	6.50	4.70	2.86
N_{signal}	38.7 ± 0.3 ± 1.3	24.2 ± 0.2 ± 0.8	17.7 ± 0.2 ± 0.7	11.7 ± 0.2 ± 0.5

Table 5.2: *Remaining events after the preselection and after each successive selection cut. The signal efficiencies (cumulated) for leptoquark decays into the $\mu\not{E}_T jj$ and $\mu\mu jj$ states are given. The assumed leptoquark mass is 200 GeV. First uncertainties are statistical, second are systematical. The systematic uncertainty on to the integrated luminosity is not included in the given number. The branching fraction β assumed to be 0.5.*

5.1.3 Multivariate analysis

The multivariate analysis has been performed for each value of β from 0.1 to 1 (with a step of 0.1). In the following, we describe in detail the event selection assuming $\beta = 0.5$.

The k-Nearest Neighbors algorithm

The optimization of the presented analysis has been performed with the use of a multivariate discriminant based on the k-Nearest Neighbors (*kNN*) algorithm. The implementation provided in TMVA [66] (Toolkit for Multivariate Analysis) has been utilized.

The k-Nearest neighbors method relies on the comparison of a test event to reference events taken from training data sets. The implemented algorithm can be interpreted as a generalization of the classical projective likelihood method [66] to n dimensions, where n is the number of variables used for the discrimination of signal against background. During the training phase, the classification of an event as being either of signal or background type is achieved by estimating the local signal-like probability density. The output of the discriminant is a variable parametrized in n dimensions with the input variables, and taking values between 0 and 1, 0 referring to the most background-like events, and 1 to the most signal-like.

The local signal-like probability density results from the counting of background and signal type nearest neighbor events within a specific volume. The distance between the test event and a neighbor event is defined with the Euclidean metric and is defined by the following equation:

$$R = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \tag{5.3}$$

where:

$\begin{cases} n \text{ is the number of input variables included in the algorithm} \\ x_i \text{ is the coordinate of the test event in the space formed by the n variables} \\ y_i \text{ is the coordinate of a neighbor event in the space formed by the n variables} \end{cases}$

The size of the volume is fixed by the number of allowed neighbors. The *kNN* algorithm is thus intrinsically adaptive, since the volume size is adjusted for the classification of each trained event. The local signal-like probability density is given by the following formula:

$$P_{signal} = \frac{N_{signal}}{k} = \frac{N_{signal}}{N_{signal} + N_{background}} \tag{5.4}$$

where:

$\begin{cases} k \text{ is the number of neighbor events with respect to the test event} \\ N_{signal} \text{ is the number of signal events among the neighbor events} \\ N_{background} \text{ is the number of background events among the neighbor events} \end{cases}$

The training phase depends on the number of input variables considered, and the number of neighbor events (k) that are requested. Variables which provide a good separation between signal and background events are preferred, but the *kNN* algorithm is not designed to include a large number of variables ($n \gtrsim 10$) [66], as the neighborhood size would increase such that the speed of the classification would suffer significantly. The value chosen for k has a direct

impact on the performance of the algorithm. A large value will tend not to describe the local behavior of the probability density function, whereas a small value will enhance its sensibility to statistical fluctuations. Reference [66] suggests values of k between 10 and 100. When using large values of k, polynomial kernel weights can be used to re-weight the probability density function according to the distance between the tested event and its neighbors. This alternative has not been considered in the analysis presented since it did not improve the classifier performance.

To sum things up, given a test event, the *kNN* algorithm response can be interpreted as a counting experiment within a n-dimensional event volume. The size of this volume is adapted according to the event density in the vicinity of the tested event, such that the number of neighbor events remains constant and equal to k.

Description of the training phase

The training phase of the *kNN* classifier has been performed on signal and simulated background samples selected with tighter cuts as in the preselection defined in Section 5.1.1. The training can thus be focused in the kinematic region of interest, i.e. closer to signal-like events and beyond the resonance of the W boson. The transverse mass $M_T(\mu, \not{E}_T)$ has been required to be greater than 110 GeV and a lower threshold on the scalar transverse variable S_T has been set to 200 GeV. Table 5.3 provides the number of remaining events in all samples after the preselection plus the preceding cuts. The new preselection is referred to as the "tight preselection".

Samples	$M_T(\mu\not{E}_T)$ >110 GeV	$S_T > 200$ GeV
$W(+jets) \to l\nu + jets$	425 ± 6 ± 54	359 ± 5 ± 45
$Z/\gamma^*(+jets) \to \mu\mu + jets$	61 ± 1 ± 7	53 ± 1 ± 7
$Z/\gamma^*(+jets) \to \tau\tau + jets$	1.9 ± 0.4 ± 0.2	1.7 ± 0.4 ± 0.2
$t\bar{t}$ (inclusive)	40.1 ± 0.4 ± 7.5	39.7 ± 0.4 ± 7.4
QCD multijet	8.8 ± 0.3 ± 1.8	8.4 ± 0.3 ± 1.7
Total Background	537 ± 6 ± 64	461 ± 5 ± 55
Data	486	422
$\epsilon_{Signal}^{\mu\nu qq}$ (%)	17.0	17.0
$\epsilon_{Signal}^{\mu\mu qq}$ (%)	7.59	7.59
N_{signal}	28.0 ± 0.3 ± 1.0	28.0 ± 0.2 ± 1.0

Table 5.3: *Remaining events after each successive additional cut of the tight preselection. The signal efficiencies for leptoquark decays into $\mu\not{E}_T jj$ and $\mu\mu jj$ final states are given. The assumed leptoquark mass is 200 GeV. First errors are statistical, second are systematical. The systematic uncertainty on the integrated luminosity is not included.*

Since each signal event considered in this channel results from the mixture of $\mu\nu qq$ and $\mu\mu qq$ generated events which are both weighted by different branching ratios ($2\beta(1-\beta)$ and β^2, respectively), the *kNN* optimization has to be performed for each considered β value separately. However, events generated in the $\mu\mu qq$ state with one non-reconstructed muon present the same topology has $\mu\nu qq$ events. As a result, the training phase will not depend on β. In the following, we quote results obtained with a training computed with $\beta = 0.5$. The case where $\beta = 0$ corresponds to a branching fraction equal to 0 for both types of generated events and is therefore not studied. The best sensitivity is obtained when repeating the training for each assumed leptoquark

mass, which has been done in the analysis presented.

Two parameters have to be tuned in order to efficiently classify events according to the *kNN* algorithm, namely the number of nearest neighbors and the number of physical variables to consider.

First, as a consistency check, the performance of the algorithm has been tested when using exactly the same variables as in the cut-based analysis (i.e., $M_T(\mu, \not{E}_T)$, S_T, and $M_T(\not{E}_T, jet_1)$). The signal efficiency after all the cut-based analysis cuts, and with respect to the tighter preselection of Table 5.3, is found to be of 43%. The corresponding background rejection is 98.1%. When varying the number of nearest neighbors from 20 to 400, background rejections increasing from 97.7 to 98.3% have been determined with the *kNN* classifier, when requiring a signal efficiency of 43%. It can thus be concluded that in the chosen configuration, the *kNN* algorithm has a similar performance as the cut-based selection.

In order to improve the selection based on the *kNN* algorithm, three additional input variables have been added, namely the transverse mass constructed from the four-vector of the second leading jet and the missing transverse energy ($M_T(\not{E}_T, jet_2)$), the invariant mass constructed from the four-vectors of the muon and the leading jet ($M(\mu, jet_1)$), and the invariant mass constructed from the four-vectors of the muon and the second leading jet ($M(\mu, jet_2)$). These three variables are contributing to the reconstructed leptoquark mass, which motivates their use. A significant improvement has been noticed when training the classifier with six variables. When varying the number of nearest neighbors from 20 to 400, background rejections increasing from 99.0 to 99.2% have been estimated with the *kNN* method, for a signal efficiency of 43%.

In the following, the number of neighbors events has been fixed to 50 for all trainings, since larger numbers did not improve the performance while exposing the optimization to potential overtraining. For the same reason, no additional input variable has been added, which would have furthermore significantly increased the processing time.

The result of the training consists in weights that can be applied to the analysis samples, and which return the *kNN* output on an event basis, given the value of the input variables. When applying these weights to the analysis samples, all the events of the signal and background samples have been considered.

So as to avoid overtraining, the training phase has been performed on half of each signal and background sample. When looping on the signal and background samples to train the *kNN* classifier, alternatively one event was kept for the training and the next one was included in a test sample. Figure 5.6 shows the comparison between the *kNN* output of the trained sample and that of the test sample, for both signal and background, when using three or six input variables and 50 neighbor events, as described in the following. Good agreement is found between the test and training samples, and thus no overtraining is observed. In addition, the figure clearly shows the increase of sensitivity when using six input variables instead of three, since the separation between signal and background is improved. Each time the *kNN* classifier was used in the work presented in this note, the training and test samples have been compared and good agreement has been observed.

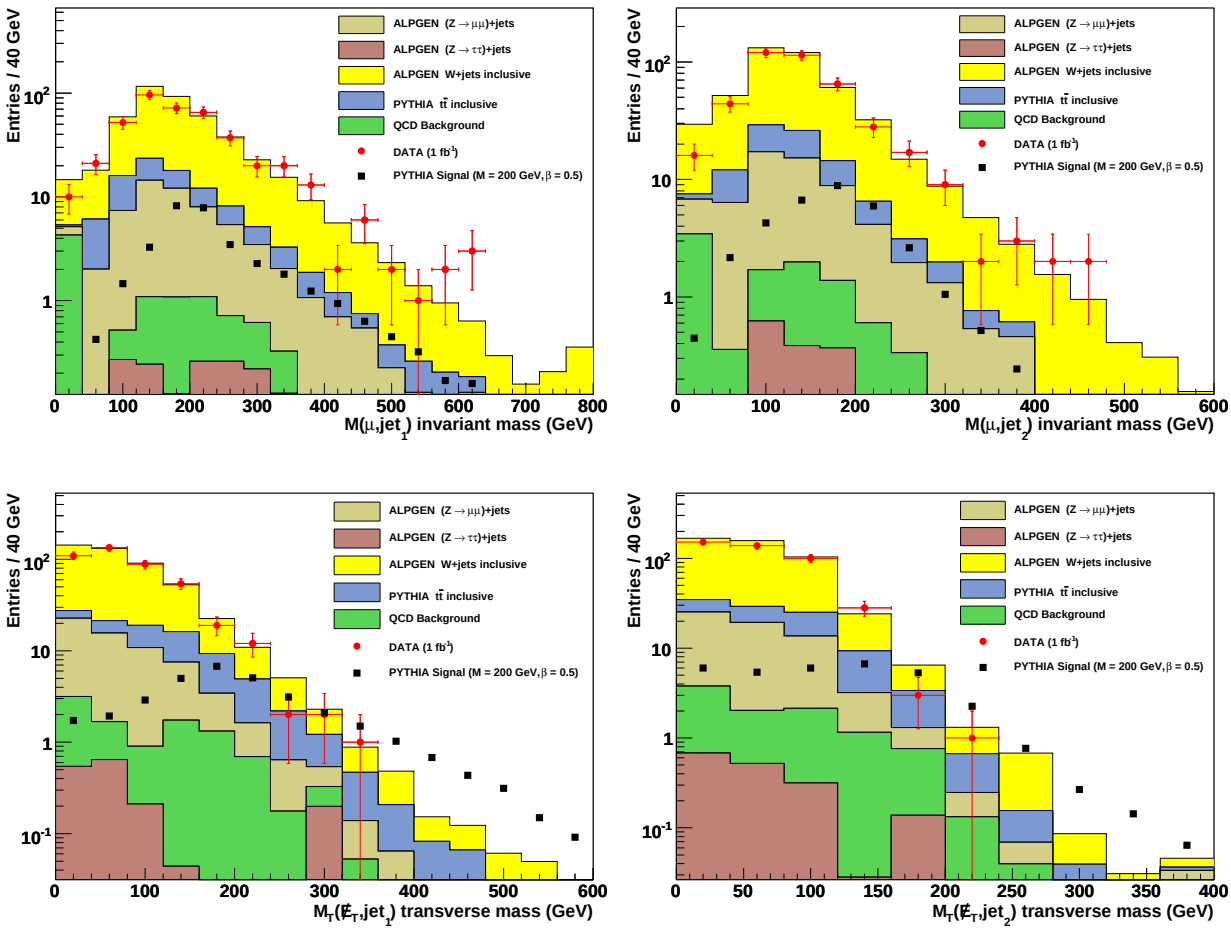

Figure 5.5: *Muon − jet invariant masses constructed from the four vectors of the muon and of one of the two leading jets (upper plots), and $\not{E}_T$ − jet transverse masses, constructed from the four vector of one of the two leading jets and the missing transverse energy (lower plots). These four variables are contributing to the reconstruction of the LQ mass. The distributions are represented after the application of the tight preselection cuts, for data (red dots), background (histograms), and signal (black squares), assuming $M_{LQ} = 200$ and $\beta = 0.5$.*

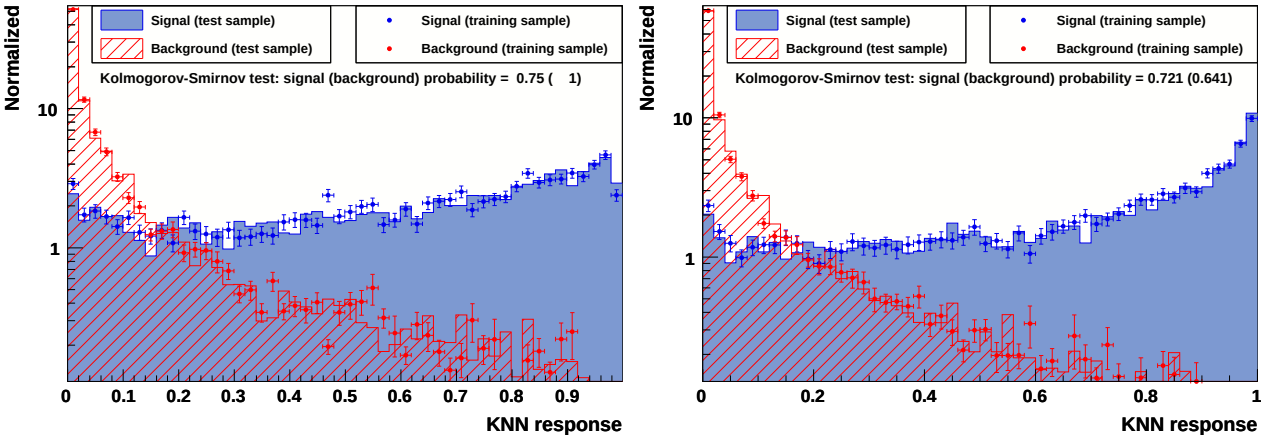

Figure 5.6: *Comparison of the kNN output variable in the training and the test samples, assuming* $n = 50$*, and 3 (left) or 6 (right) input variables,* $M_T(\mu, \not{E}_T)$ *and* S_T*, or* $M_T(\mu, \not{E}_T)$*,* S_T*,* $M(\mu, jet_1)$*,* $M(\mu, jet_2)$*,* $M_T(\not{E}_T, jet_1)$ *and* $M_T(\not{E}_T, jet_2)$*, respectively.*

Selected events and signal efficiencies

Only the case where $M_{LQ} = 200\,\text{GeV}$ and $\beta = 0.5$ is presented in detail, in order to be able to compare the results with the cut-based analysis.

The output of the *kNN* variable for data, background and signal ($M_{LQ} = 200\,\text{GeV}$, $\beta = 0.5$) after the tight preselection is represented on the left plot of Figure 5.7. The right plot of the same figure shows the expected leptoquark pair production upper cross section limit as function of the lower threshold of a cut placed on the *kNN* output variable. In comparison with the optimization of the cut-based analysis (see Section 5.1.2), a significantly lower expected upper cross section limit is found, namely 0.13 pb (when imposing *kNN output*> 0.75) against 0.18 pb for the cut-based analysis. Yet, instead of simply cutting on the *kNN* variable to perform the event selection, another strategy has been chosen to optimize the sensitivity. In order to keep the full information about the shape of the *kNN* output both at low and high signal efficiency, the *kNN* distribution has been divided into 6 bins of variable size which is decreasing with increasing signal efficiencies. The same binning has been used for the other assumed leptoquark masses and the other β values. Results relative to the limit calculations based on bins of the *kNN* output variable are discussed in Section 6.2.2).

The edges and content of each *kNN* bin, including systematic uncertainties, are detailed in Table 5.4, assuming $M_{LQ} = 200\,\text{GeV}$ and $\beta = 0.5$. The signal efficiencies in each bins, for all the assumed leptoquark masses, and for the case $\beta = 0.5$, are provided in Table 5.5.

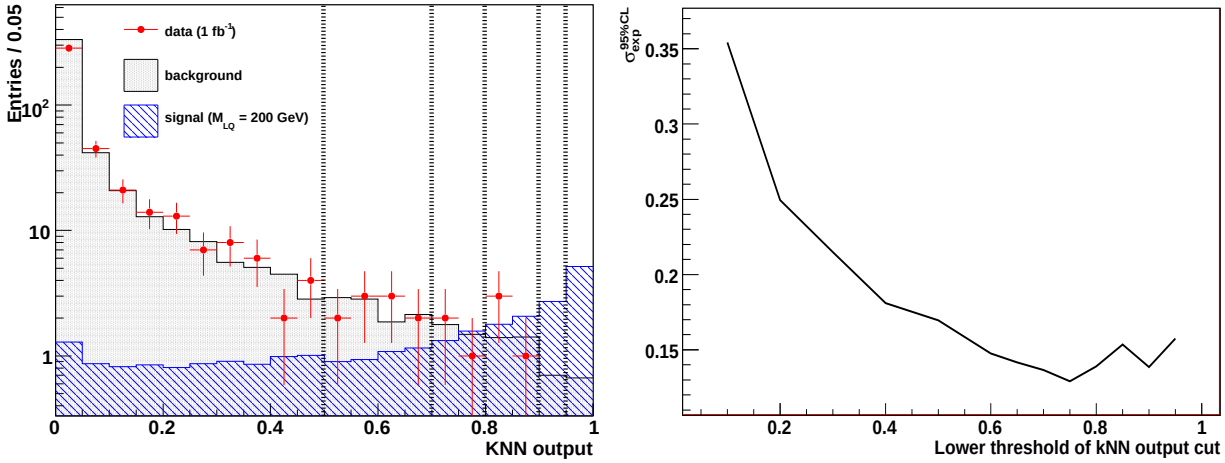

Figure 5.7: *Left: distribution of the kNN output variable for data (red dots), background (black histogram), and signal (blue histogram) events. The chosen bins are separated by the dashed vertical black lines. Right: Expected (95% CL) upper cross section limit on the leptoquark pair production as function of the lower threshold of the one-dimensional cut on the kNN output variable.*

Samples	Bin 1: $0 < kNN < 0.5$	Bin 2: $0.5 < kNN < 0.7$	Bin 3: $0.7 < kNN < 0.8$
$W(+jets) \to l\nu + jets$	348 ± 5 ± 44	6.1 ± 0.5 ± 1.3	1.8 ± 0.3 ± 0.4
$Z/\gamma^*(+jets) \to \mu\mu + jets$	51 ± 1 ± 6	0.93 ± 0.11 ± 0.15	0.39 ± 0.07 ± 0.08
$Z/\gamma^*(+jets) \to \tau\tau + jets$	1.5 ± 0.3 ± 0.2	0.20 ± 0.19 ± 0.03	0.0037 ± 0.0014 ± 0.0007
$t\bar{t}$ (inclusive)	34.9 ± 0.4 ± 6.5	2.5 ± 0.1 ± 0.5	1.0 ± 0.1 ± 0.2
QCD multijet	8.3 ± 0.3 ± 1.7	0.030 ± 0.017 ± 0.006	0.020 ± 0.014 ± 0.004
Total Background	444 ± 5 ± 53	9.8 ± 0.6 ± 1.7	3.3 ± 0.3 ± 0.6
Data	405	10	3
$\epsilon^{\mu\nu qq}_{Signal}$ (%)	5.41	2.45	1.80
$\epsilon^{\mu\mu qq}_{Signal}$ (%)	3.72	1.12	0.539
N_{signal}	9.3 ± 0.1 ± 0.4	4.1 ± 0.1 ± 0.1	2.9 ± 0.1 ± 0.2

Samples	Bin 4: $0.8 < kNN < 0.9$	Bin 5: $0.9 < kNN < 0.95$	Bin 6: $0.95 < kNN < 1$
$W(+jets) \to l\nu + jets$	1.7 ± 0.3 ± 0.5	0.31 ± 0.09 ± 0.24	0.39 ± 0.13 ± 0.18
$Z/\gamma^*(+jets) \to \mu\mu + jets$	0.23 ± 0.04 ± 0.01	0.11 ± 0.02 ± 0.01	0.064 ± 0.006 ± 0.022
$Z/\gamma^*(+jets) \to \tau\tau + jets$	0.0028 ± 0.0007 ± 0.0002	0.0031 ± 0.0014 ± 0.0004	0.0020 ± 0.0008 ± 0.0007
$t\bar{t}$ (inclusive)	0.84 ± 0.05 ± 0.19	0.26 ± 0.03 ± 0.05	0.21 ± 0.03 ± 0.06
QCD multijet	0.030 ± 0.017 ± 0.006	0.020 ± 0.014 ± 0.004	0.0099 ± 0.0099 ± 0.0020
Total Background	2.8 ± 0.3 ± 0.5	0.70 ± 0.10 ± 0.25	0.67 ± 0.14 ± 0.22
Data	4	0	0
$\epsilon^{\mu\nu qq}_{Signal}$ (%)	2.38	1.71	3.28
$\epsilon^{\mu\mu qq}_{Signal}$ (%)	0.748	0.463	0.625
N_{signal}	3.9 ± 0.1 ± 0.1	2.7 ± 0.1 ± 0.1	5.2 ± 0.1 ± 0.3

Table 5.4: *Content of each bin of the kNN variable after the tight preselection . The signal efficiencies (cumulative) for leptoquark decays into the $\mu\nu qq$ and $\mu\mu qq$ states are given, as well as the number of events in the data and background samples. The assumed leptoquark mass is 200 GeV, and β is taken equal to 0.5. First uncertainties are statistical, second are systematical. The systematic uncertainty on the integrated luminosity is not included.*

	Bin 1		Bin 2		Bin 3		Bin 4		Bin 5		Bin 6	
M_{LQ} (GeV)	$\epsilon^{\mu\nu qq}_{Signal}$	$\epsilon^{\mu\mu qq}_{Signal}$	$\epsilon^{\mu\nu qq}_{Signal}$	$\epsilon^{\mu\mu qq}_{Signal}$	$\epsilon^{\mu\nu qq}_{Signal}$	$\epsilon^{\mu\mu qq}_{Signal}$	$\epsilon^{\mu\nu qq}_{Signal}$	$\epsilon^{\mu\mu qq}_{Signal}$	$\epsilon^{\mu\nu qq}_{Signal}$	$\epsilon^{\mu\mu qq}_{Signal}$	$\epsilon^{\mu\nu qq}_{Signal}$	$\epsilon^{\mu\mu qq}_{Signal}$
140	6.693	2.729	1.731	0.553	1.033	0.215	0.979	0.205	0.261	0.046	0.053	0.001
160	6.389	3.65	2.339	0.703	1.193	0.303	1.685	0.338	1.022	0.176	0.766	0.089
180	6.349	3.933	2.64	0.955	1.544	0.531	2.123	0.552	1.27	0.224	1.389	0.21
200	5.407	2.974	2.445	1.177	1.796	0.725	2.385	0.971	1.709	0.618	3.279	1.117
220	4.633	3.629	2.416	1.207	1.827	0.707	2.585	0.821	1.784	0.525	5.188	1.116
240	3.658	2.153	2.228	1.137	1.626	0.814	2.59	1.122	2.045	0.969	7.499	2.643
260	3.284	2.21	1.938	1.022	1.435	0.78	2.323	1.039	1.937	0.868	9.167	3.263
280	2.571	1.695	1.498	0.997	1.342	0.628	2.069	1.055	1.936	0.849	11.417	4.284
300	2.116	1.498	1.326	0.738	1.232	0.612	1.942	1.037	1.829	0.883	13.009	5.113
320	1.923	1.477	1.116	0.662	1.045	0.524	1.644	0.938	1.715	0.709	14.563	5.693

Table 5.5: *Signal efficiencies (in % and cumulative) after all cuts, for leptoquark decay into the $\mu\nu qq$ and $\mu\mu qq$ final states, and for all the considered assumed leptoquark masses. The branching fraction β is taken equal to 0.5.*

5.1.4 Systematic uncertainties

The systematic uncertainties on signal and simulated backgrounds are assumed to be Gaussian and symmetric. For clarity, we only quote the errors on the cut-based and the multivariate analysis. When not explicitly mentioned, the quoted uncertainties relative to the multivariate analysis refers to the case ($\beta = 0.5$, $M_{LQ} = 200\,\text{GeV}$). Since the training of the *kNN* has been performed for each (M_{LQ},β) case, the respective errors on both the signal and background may vary. Thus, for errors that are not treated as constant, the computation has been performed for each (M_{LQ},β) point of the analysis, and after each selection cut.

In the following, we describe the different uncertainties that have been considered in the analysis.

Integrated luminosity

The relative uncertainty on the integrated luminosity is equal to 6.1% [67], and is treated as a global constant error.

Muon acceptance and smearing

For both signal and background, relative uncertainties of 0.7%, 0.8% and 1.7% are included for the muon identification, the track identification, and the isolation working point, respectively (following Reference [63]). An relative uncertainty of 0.8%, based on the results of Section 4.3.2, has been derived for the combination of the muon triggers. These errors are treated as global constant errors.

The uncertainty due to the muon p_T oversmearing is estimated by shifting the Gaussian central prediction by one standard deviation in each direction, and for each considered set of selection cuts. Errors are then symmetrized in a conservative way.

For the cut-based analysis, after applying all selection cuts, the relative uncertainty due to the muon p_T oversmearing is found to be of 14.6%, 6.9% and 1.6% for W/Z backgrounds, the $t\bar{t}$ production, and the signal, respectively.

The large values, obtained for the W and the Z boson production in high p_T regions, originates from the extrapolation of the smearing parameters that are derived at much lower energies, namely in the regions of Z boson and J/ψ production.

The multivariate analysis includes relative uncertainties on the muon p_T oversmearing varying from 0.8% to 3.2%, in the case of the signal prediction. For the W/Z backgrounds, this error can be as low as 5.7% and take values up to 27%. Table 5.6 lists the uncertainties for the 6 bins of the multivariate analysis. Anti-correlations between bins have been taken into account, and are represented by a difference in sign.

It is worth mentioning that for other (M_{LQ},β) analysis points, slightly different errors may be calculated. Higher values, especially for background, arise from statistical fluctuations in the *kNN* bins, and can be as high as 64%.

All the muon related uncertainties are added in quadrature.

Sample	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
W/Z	12	13	16	-5.7	5.7	27
$t\bar{t}$	4.0	6.6	7.5	13	4.7	23
Signal	-1.0	-1.5	-3.2	-0.8	1.0	-1.8

Table 5.6: *Muon transverse momentum smearing uncertainty (in %) on background and signal, in the case where $M_{LQ} = 200\,GeV$ and $\beta = 0.5$. Anti-correlations between bins are represented by a difference of sign.*

The jet shifting, smearing and identification

The relative uncertainties on the jet shifting, smearing and identification are estimated by shifting the Gaussian central prediction by one standard deviation (positively and negatively). The calculation is done at each level of the event selection, and errors are symmetrized conservatively.

For the cut-based analysis, after all cuts, the relative error on the jet shifting is of 3.6%, 5.0% and 2.7%, for the W/Z backgrounds, the $t\bar{t}$ production, and the signal sample, respectively. The relative uncertainty on the jet smearing is found to be very low for each simulated sample. Values from 0.03% to 0.13% are found. The relative error of the jet identification varies from 0.26% to 0.56%, which is due to the fact that the jet reconstruction efficency is on its plateau for highly energetic jets.

In the case of the multivariate analysis, relative uncertainties on the jet shifting (reported in Figure 5.7) are varying from 0.63% to 20%, from 0.011% to 11%, and from 0.052% to 3.2%, for the W/Z production, the $t\bar{t}$ background, and the signal prediction, respectively. The derived relative errors on the jet smearing mainly stay under 1%, varying from negligible values to 1.4% among all the simulated samples. As regards the background, the relative error on the jet identification varies from 0.037% to 1.2% for the five first *kNN* bins. Relative errors on up to 4.9% are derived for the most signal-like bin (the last one), because of statistical fluctuations. In the case of the signal, the relative uncertainty in the jet identification remains very low, and varies from 0.29% to 0.62%.

For all the jet uncertainties, higher values are sometimes found in other (M_{LQ}, β) analysis points, due to high statistical fluctuations. For instance, a relative error of 57% for the jet shifting has been derived for the point ($M_{LQ} = 240\,\mathrm{GeV}$, $\beta = 1$), in the case where the W/Z background contributes with only 0.21 event, with a statistical error of 0.11% (in the most signal-like *kNN* bin). It it thus likely to be a statistical artifact.

Sample	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
W/Z	3.8	10.0	12	0.63	9.5	-20
$t\bar{t}$	0.011	3.7	11	5.7	8.1	7.6
Signal	-2.1	0.76	3.0	-0.052	3.6	3.2

Table 5.7: *Jet shifting uncertainty (in %) on background and signal, in the case of the multivariate analysis ($M_{LQ} = 200\,GeV$ and $\beta = 0.5$), and as function of the kNN bins. Anti-correlations between bins are represented by a difference in sign.*

All the jet uncertainties are added in quadrature.

The ALPGEN predicted jet p_T shape in the W sample

An additional systematic error is evaluated to account for uncertainties in the correct modeling of the jet p_T shape in W events. This number has been evaluated by comparing the second jet p_T distribution observed in data with the predictions of ALPGEN and PYTHIA, in a kinematic region dominated by W production. The leading jet p_T has not been considered since it is well described in both ALPGEN and PYTHIA.

The difference between ALPGEN and PYTHIA can be significant for the second leading jet p_T. Therefore, the method applied in this analysis consisted in mixing the ALPGEN and PYTHIA sample up to the extent where the Kolmogorov-Smirnov test between the mixture and real data is still acceptable. The applied test providing a probability of consistence, the lower output value to be accepted has been chosen to be 31.73%, which refers to a standard deviation of one sigma.

A mixture composed of up to 30% of PYTHIA has been found to be acceptable. Figure 5.8 provides the distribution of the second leading jet p_T for ALPGEN, PYTHIA and data, as well as the Kolmogorov-Smirnov test behavior.

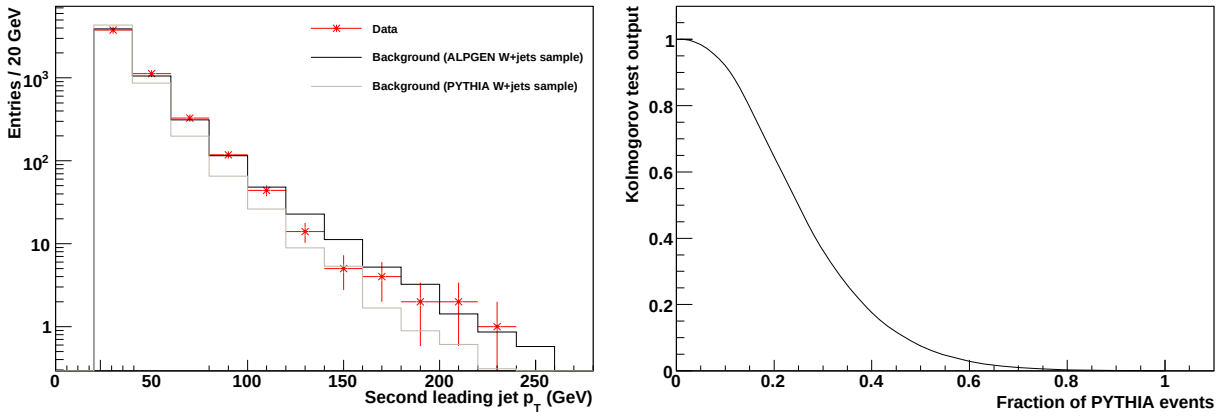

Figure 5.8: *Left: distribution of the second leading transverse momentum in data and background events, when simulating the W boson production with* ALPGEN *(black curve) or* PYTHIA *(gray curve).Right: Kolmogorov test output describing the compatibility between the data and background second leading transverse momentum distribution, as function of the fraction* PYTHIA *events included in addition to* ALPGEN *events for the description of the W boson production.*

In the cut-based analysis, the defined mixture of W events differs by 15% from the ALPGEN prediction.

The multivariate analysis includes various values for the jet p_T shape error, depending on the considered *kNN* bins. The background-like bin (the first one) have the lowest error, namely 2.1%, Other bins are subject to larger errors, from 8.9% to 30%. When the statistics of the PYTHIA sample turns out to be too small for a given bin, rather large errors have been calculated, such as 77% in the present case. Table 5.8 lists the jet p_T shape uncertainties in the case where $M_{LQ} = 200\,\text{GeV}$ and $\beta = 0.5$.

Sample	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
jet p_T shape uncertainty	-2.1	13.9	8.9	30	-77	30

Table 5.8: *Uncertainty (in %) due to the* ALPGEN *jet p_T shape in the case of the multivariate analysis, as function of the kNN bins, and assuming $M_{LQ} = 200\,GeV$ and $\beta = 0.5$. Anti-correlations between bins are represented by a difference in sign.*

The normalization of W/Z samples

An error of 1.2% on the W/Z backgrounds is added to account for uncertainties in its normalization to data. This uncertainty has been estimated by varying the normalization window from]50, 110[GeV to]70, 90[GeV, and is treated as a global constant error.

The QCD multijet background

The uncertainty on the QCD multijet background is estimated to be 20%, which includes both the uncertainty on the normalization and on the extrapolation of the QCD templates, which are defined using anti-isolation cuts for the reconstructed muon, to the signal region. The same error on the QCD background is considered at each event selection level.

The $t\bar{t}$ production cross section

A global constant uncertainty of 18% on the theoretical prediction of the $t\bar{t}$ pair production cross section has been taken into account. A more recent uncertainty can be found in Reference [68].

The gluon radiation

Changes in the gluon radiation influences jets and has thus an effect on acceptance. In order to account for the uncertainty due to the gluon radiation in the initial and final state for the signal, the amount and the strength of the showering radiations have been varied according to Reference [69]. This corresponds to the parameters of PYTHIA which describe the maximum virtuality scale in the parton showering (Q^2_{max}), and Λ_{QCD}.

The variations of Q^2_{max} have been described by comparing two signal samples, which we refer to as *sample Q1* and *sample Q2*. In *sample Q1*, the PYTHIA parameters PARP(67) and PARP(71), of central values 4, have been shifted down to 0.25 and 1, respectively, while in *sample Q2*, they have been set to 4 and 16, respectively.

Λ_{QCD} has been tuned by comparing two other samples: *sample L1* and *sample L2*. The PYTHIA parameters PARP(61), PARP(72) and PARJ(81) have been symmetrically shifted down to 0.10 and up to 0.23, in *sample L1* and *sample L2*, respectively.

The four samples have been compared to the central signal sample. On the one hand, the signal rate in *sample Q1* differs from the central value by 0.5%, while a deviation of 0.2% is observed in *sample Q2*. On the other hand, *sample L1* and *sample L2* show a shift of 0.2% and 1%, respectively. To be conservative a global constant uncertainty of 1.5% due to the modeling of gluon radiation in the initial and final state radiation has been used.

The parton density functions

The uncertainty on the signal acceptance due to uncertainties in the parton density functions (PDF) can be estimated in the signal samples by comparing the central signal rate expectation with the rate reweighted according to the factor below [70]:

$$\frac{f_a^i(x_a, Q) \times f_b^i(x_b, Q)}{f_a^0(x_a, Q) \times f_b^0(x_b, Q)} \tag{5.5}$$

where a and b refer to the partons in the initial state, f^0 is the PDF used for the signal generation (CTEQ6L1) and the f^i are the CTEQ6.1M error PDF sets.

40 error functions are included in the reweighting. An error on the signal acceptance of 1.6% has been derived. The same error has been used for all the signal samples, as it did not significantly vary, and remained under 1.6%. This uncertainty is treated as constant with respect to the event selection level.

5.2 Selection in the $\mu\mu jj$ final state

5.2.1 Preselection

Preselection cuts

Events are required to have at least two loose muons found within $|\eta| < 2$, with loose isolations. The selected muons are in addition required to be matched to a central track of at least medium quality, with a transverse momentum exceeding 20 GeV. At least two jets are required to deposit energy in the region $|\eta| < 2.5$ of the calorimeter, with transverse energies greater than 25 GeV.

The $M(\mu, \mu)$ invariant mass reconstructed from the four-vector of the two muons is required to exceed 50 GeV, which reduces the QCD multijet contribution to a negligible extent. In addition, events with $M(\mu, \mu)$ lower than 50 GeV are not in the signal region of the presented search, which is beyond the Z boson resonance.

All the reconstruction cuts plus the precedent cuts will be referred to as preselection cuts. Once applied, 913 data events and 930 background events are remaining, as listed in Table 5.10. The signal efficiency, as defined in Section 5.1.1, is 41.4% at preselection level.

The contribution of events arising from WW and WZ and ZZ bosons production is not taken into account since negligible [35].

Background rescaling

The $Z/\gamma^* \to \mu\mu$ and $Z/\gamma^* \to \tau\tau$ samples are simultaneously normalized to data by the factor $F_{Z/W}$, which is determined at preselection level in the region of the Z boson peak ($M(\mu, \mu) \in$]60, 120[GeV). As in the $\mu \not{E}_T jj$ channel, we normalize the contribution of these backgrounds by the same factor because they present similar topologies. The contributions of the $t\bar{t}$ and W boson production in the Z boson resonance region is negligible.

A scaling factor $F_{Z/W} = 1.28$ has been derived, all simulated backgrounds being initially normalized to data with the use of NNLO cross section predictions (as described in Section 3.1.2).

Correction of the $M(\mu\mu)$ invariant mass due to the potential mismeasurements of muons transverse momenta

The mismeasurement of the muon transverse momentum, resulting in an overestimated di-muon invariant mass, could be circumvented by cutting on the azimuthal angle between the leading muon and the missing transverse energy, as was done for $\mu \not{E}_T jj$ channel. This method has the drawback of rejecting both signal and Standard Model background events. Nevertheless, no missing transverse energy is expected in $\mu\mu jj$ events, as they do not include neutrinos, and since missing transverse energy arising from QCD multijet production is discarded by the preselection cuts. As a consequence, the vectorial sum of all the muons and jets transverse momenta of a given event is expected to vanish, which make it possible to balance the muon transverse momentum.

Figure 5.9 provides the distribution of the di-muon invariant mass at preselection level. Since there is a good agreement between the data and the Standard Model prediction, it turns out that the overestimation of the muon transverse momentum resolution is adequately described in simulation.

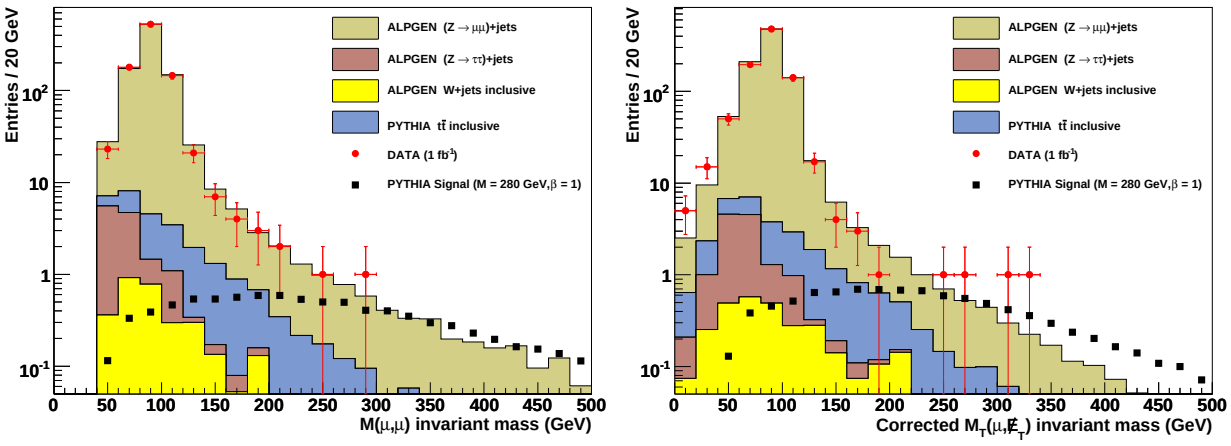

Figure 5.9: *Distributions of the uncorrected (left) and corrected (right) di-muon invariant masses for data (red dots), background (histograms), and signal (black boxes). For the signal distribution, $M_{LQ} = 280\ GeV$ and $\beta = 1$ are assumed. The di-muon invariant mass correction is obtained by balancing the leading muon transverse momentum.*

As motivated in Section 5.2.2, the di-muon invariant mass constitutes an efficient variable to cut on to discriminate signal from the Standard Model background which mainly consists in Z/γ^* boson production. However, overestimated $M(\mu, \mu)$ values contribute to a wider tail in the $M(\mu, \mu)$ distribution, and thus reduce the separation between signal and background.

Since the missing transverse energy should be negligible in $\mu\mu jj$ events, we interpret the observed missing transverse energy as the consequence of the overestimation of the leading muon transverse momentum. The direction of the leading muon is however not affected by the mismeasurement, we therefore only correct the norm of the leading muon p_T. The correction consists in subtracting the projection of the missing transverse energy from the direction of the leading muon in the transverse plan, as formulated by the following substitution:

$$|p_T| \rightarrow |p_T| - \frac{\overrightarrow{p}_T . \overrightarrow{\not{E}}_T}{|p_T|} \tag{5.6}$$

In order to enhance the discriminating power of the $M(\mu,\mu)$ distribution, we reconstructed the di-muon invariant mass using a corrected leading muon transverse momentum. The resulting invariant mass is represented in Figure 5.9 at preselection level. To further optimize the separation between signal and background, especially in the tail of the distribution, we define the minimal di-muon invariant mass ($M(\mu,\mu)^{min}$) as being the minimum of the corrected and the original value. $M(\mu,\mu)^{min}$ is provided in Figure 5.11 at preselection level.

Comparison of the Data with the Standard Model expectation

Figure 5.10 shows the comparisons of the data and the Standard Model expectation for several distributions, after the preselection cuts. A good agreement with the Standard Model expectation is observed for the muon and jet transverse momenta. The missing transverse energy and its azimuthal angle with the leading muon transverse momentum, which are used to balance the leading muon transverse momentum, are also well described by the background.

5.2.2 Cut-based analysis

Discriminating variables and optimization

Two variables have been considered for the signal selection and background rejection, namely the minimal di-muon invariant mass $M(\mu,\mu)^{min}$, and the scalar transverse variable S_T. For this channel, we define S_T as being the scalar sum of the two leading muons and the two leading jets transverse momenta:

$$S_T = p_T^{\mu_1} + p_T^{\mu_2} + p_T^{jet_1} + p_T^{jet_2} \tag{5.7}$$

The choice of the di-muon invariant mass as discriminant is motivated by the need of removing the Z boson peak. As in the $\mu\not{E}_T jj$ channel, the S_T variable provides a good separation between signal and background, since the decay products of the leptoquark pair are more energetic than those of the background processes. The distributions of $M(\mu,\mu)^{min}$ and S_T at preselection level are provided in Figure 5.11. Both variables are well described by the Standard Model prediction, and can thus be used for the event selection.

The optimization of the cuts on $M(\mu\mu)^{min}$ and S_T has been performed with the use of the significance as discriminant, and assuming a leptoquark mass of 280 GeV. We define the significance as being the number of signal events divided by the square root of the sum of signal and background events ($\frac{S}{\sqrt{S+B}}$). The set of cuts on $M(\mu\mu)^{min}$ and S_T which maximizes the significance has been chosen. First the cut value on $M(\mu\mu)^{min}$ has been set, and then the cut value on S_T. Several iterations have been done until the values of the cuts remain unchanged. The best set of rectangular cuts, which maximizes the significance, corresponds to:

$$\begin{cases} M(\mu,\mu)^{min} > 120 \text{ GeV} \\ S_T > 420 \text{ GeV} \end{cases}$$

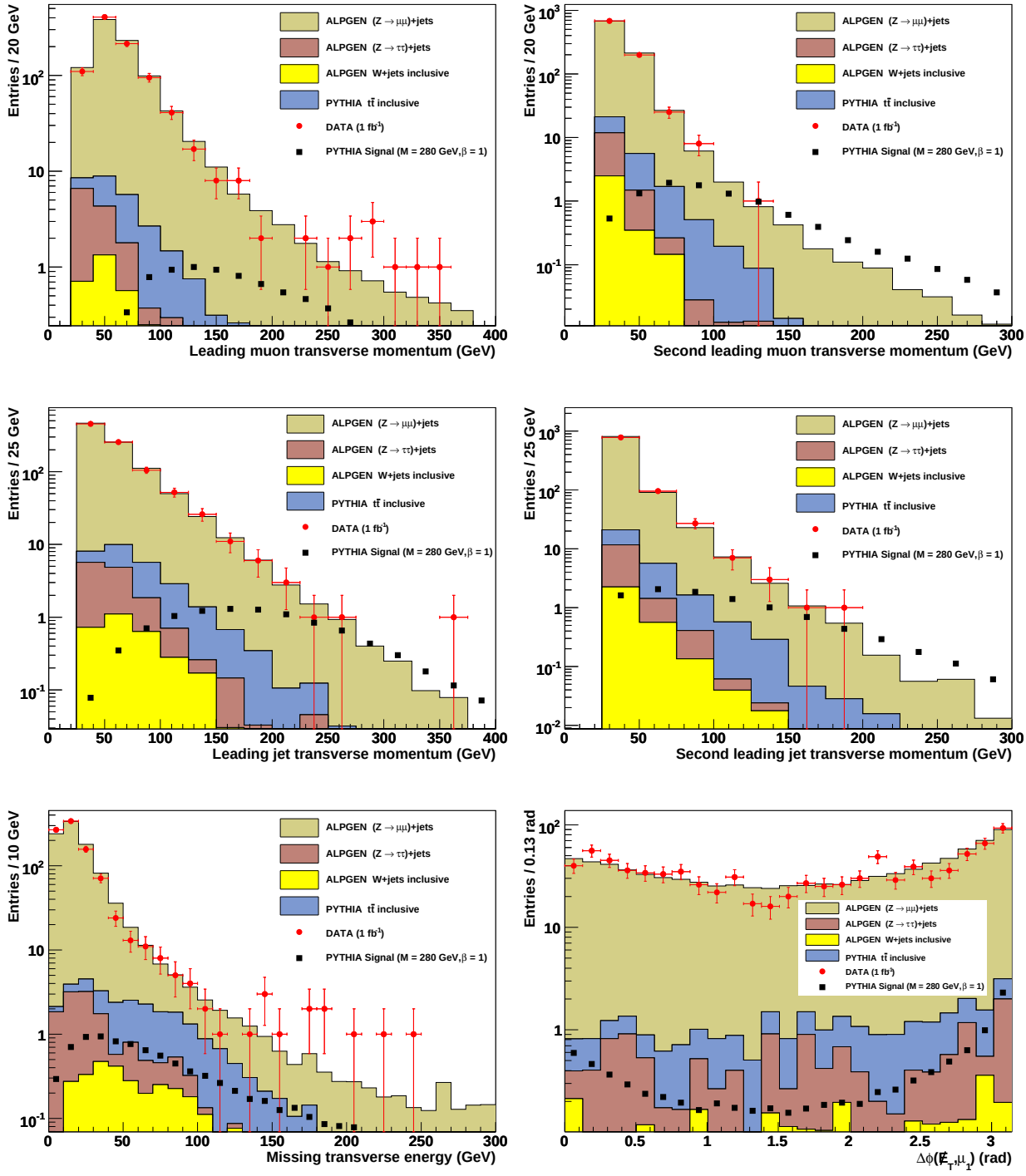

Figure 5.10: *Comparison between the data (red dots) and the Standard Model expectation (colored histograms) for several distributions: the transverse momentum of the two leading muons and jets, the missing transverse energy, and the azimuthal angle between the leading muon and the missing transverse energy. The black squares describe the leptoquark signal, generated with* $M_{LQ,gen} = 280$ *GeV and assuming* $\beta = Br(LQ \to \mu q) = 1$.

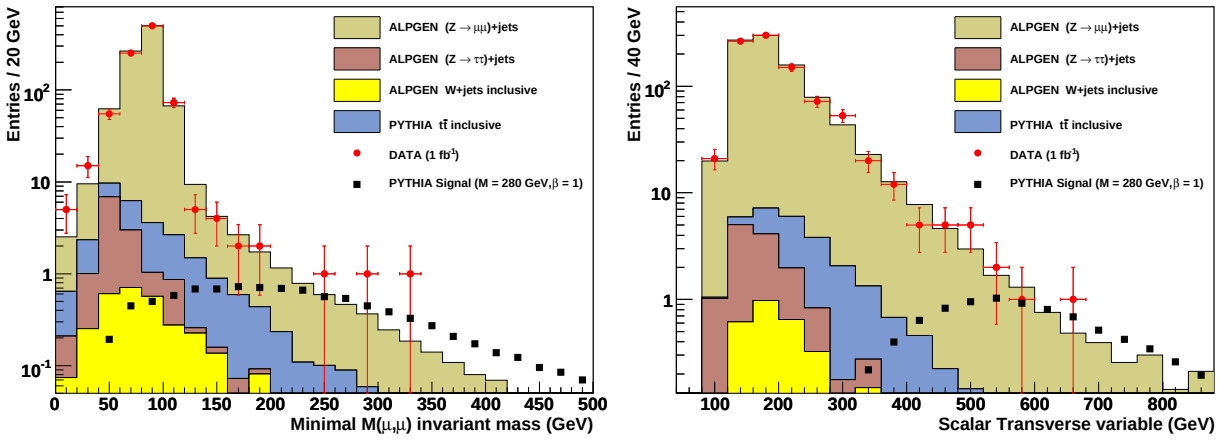

Figure 5.11: *Distributions of the cut-based analysis selection variables (namely $M_T(\mu,\mu)^{min}$ and S_T) at preselection level. The data (red dots), background (histograms), and signal (black squares) are represented. For the signal distribution, $M_{LQ} = 280\,GeV$ and $\beta = 1$ are assumed.*

Selected events and signal efficiencies

Tables 5.9 and 5.10 provide the signal efficiencies after all cuts for the various assumed leptoquark masses, and the cut flow including the preselection and the two selection cuts, respectively.

After all the cuts, events arising from the $Z \to \mu\mu$ process and $t\bar{t}$ production turn out to be the main contribution to the leptoquark signal. 2 data events and 2.9 expected background events are surviving the cut-based selection.

As in the case of the $\mu\not{E}_T jj$ channel, no evidence for leptoquark has been observed and upper limits on the leptoquark pair production cross section are calculated and interpreted in Section 6.3.1. Systematic uncertainties on the number of remaining events (appearing in Table 5.10 for the case $M_{LQ} = 280\,\text{GeV}$, $\beta = 1$) are included in the limit estimation, and detailed in Section 5.2.4. In order to further enhance the sensitivity of the search, an event selection based on a multivariate classifier has been considered, and is described in the next section.

M_{LQ} (GeV)	$\epsilon^{\mu\mu qq}_{Signal}$ (%)
140	4.84
160	7.89
180	11.9
200	16.4
220	20.6
240	24.5
260	27.7
280	30.7
300	32.6
320	33.8

Table 5.9: *Signal efficiencies (cumulative) after all cuts, for leptoquark pairs decay into the $\mu\mu qq$ final state, and for all the considered assumed leptoquark masses.*

Samples	Preselection	$M(\mu,\mu)^{min}$ >120 GeV	$S_T > 420$ GeV
$W(+jets) \to l\nu + jets$	3.0 ± 0.3 ± 0.2	0.51 ± 0.12 ± 0.04	0.020 ± 0.012 ± 0.002
$Z/\gamma^*(+jets) \to \mu\mu + jets$	901 ± 5 ± 148	18.4 ± 0.3 ± 3.1	2.3 ± 0.1 ± 0.4
$Z/\gamma^*(+jets) \to \tau\tau + jets$	11 ± 1 ± 2	0.11 ± 0.01 ± 0.02	0.034 ± 0.005 ± 0.006
$t\bar{t}$ (inclusive)	15.6 ± 0.2 ± 2.9	3.5 ± 0.1 ± 0.6	0.51 ± 0.05 ± 0.11
Total Background	930 ± 5 ± 151	22.5 ± 0.4 ± 3.3	2.9 ± 0.1 ± 0.5
Data	913	16	2
$\epsilon_{Signal}^{\mu\mu qq}$ (%)	41.4	34.0	30.7
N_{signal}	9.8 ± 0.1 ± 0.3	8.0 ± 0.1 ± 0.3	7.2 ± 0.1 ± 0.2

Table 5.10: *Remaining events after the preselection and after each selection cut. The signal efficiencies (cumulative) for leptoquark pairs decay into the μμqq final state are given. The assumed leptoquark mass is 280 GeV. First uncertainties are statistical, second are systematical. The systematic uncertainty on the integrated luminosity is not included.*

5.2.3 Multivariate analysis

The multivariate analysis has been performed for each value of β from 0.1 to 1 (with a step of 0.1). The event selection assuming $\beta = 1$ is described in detail in the following.

Training

The same training strategy as described in Section 5.1.3 has been followed (still 50 nearest neighbors have been considered). However, in the present case, the training has been performed only for $\beta = 1$. For other values of β, we use the same *kNN* weights as those obtained for $\beta = 1$. Since only the $\mu\mu qq$ decay contributes, all the signal events are weighted by the same branching fraction, namely β^2. Thus, the shape of the signal distributions is not affected by variations in β, and the shape of the *kNN* output is not modified as well.

As in Section 5.1.3, the *kNN* classifier has been trained after requiring tighter preselection cuts, which are described in Table 5.11. The original preselection of Section 5.2.1 plus the additional cuts are referred to as the tight preselection. The *kNN* has first been trained using exactly the same variables as in the cut-based analysis (i.e., $M(\mu,\mu)^{min}$ and S_T), and a similar performance as in the cut-based analysis has been obtained. However, for the same reasons as in Section 5.1.3, we divided the *kNN* output into 6 bins, and included four additional input variables to the classifier, namely the four $muon - jet$ invariant masses constructed from the four vectors of the two leading muons and the two leading jets. The choice of these variables is motivated by the fact that they contribute to the reconstructed leptoquark mass. Figure 5.12 shows their distributions once the tight preselection is applied.

Samples	$M(\mu,\mu)^{min}$ >100 GeV	$S_T > 200$ GeV
$W(+jets) \to l\nu + jets$	0.78 ± 0.14 ± 0.10	0.58 ± 0.12 ± 0.07
$Z/\gamma^*(+jets) \to \mu\mu + jets$	83 ± 1 ± 16	50.0 ± 0.8 ± 9.6
$Z/\gamma^*(+jets) \to \tau\tau + jets$	0.70 ± 0.52 ± 0.14	0.68 ± 0.52 ± 0.13
$t\bar{t}$ (inclusive)	5.3 ± 0.1 ± 1.0	4.8 ± 0.1 ± 0.9
Total Background	90 ± 1 ± 16	56 ± 1 ± 10
Data	89	59
$\epsilon^{\mu\mu qq}_{Signal}$ (%)	36.5	36.5
N_{signal}	8.6 ± 0.1 ± 0.3	8.6 ± 0.1 ± 0.3

Table 5.11: *Remaining events after the additional cuts of the tight preselection. The signal efficiencies (cumulative) for leptoquark decays into the $\mu\mu qq$ state is provided. The assumed leptoquark mass is 280 GeV. First errors are statistical, second are systematical. The systematic uncertainty on to the integrated luminosity is not included in the given number.*

Selected events and signal efficiencies

Only the case where $M_{LQ} = 280\,\text{GeV}$ and $\beta = 1$ is presented in detail, which can be directly compared to the cut-based analysis.

The output of the *kNN* variable for data, background and signal ($M_{LQ} = 280\,\text{GeV}$, $\beta = 1$) after the tight preselection is shown on the left plot of Figure 5.13. The right plot shows the

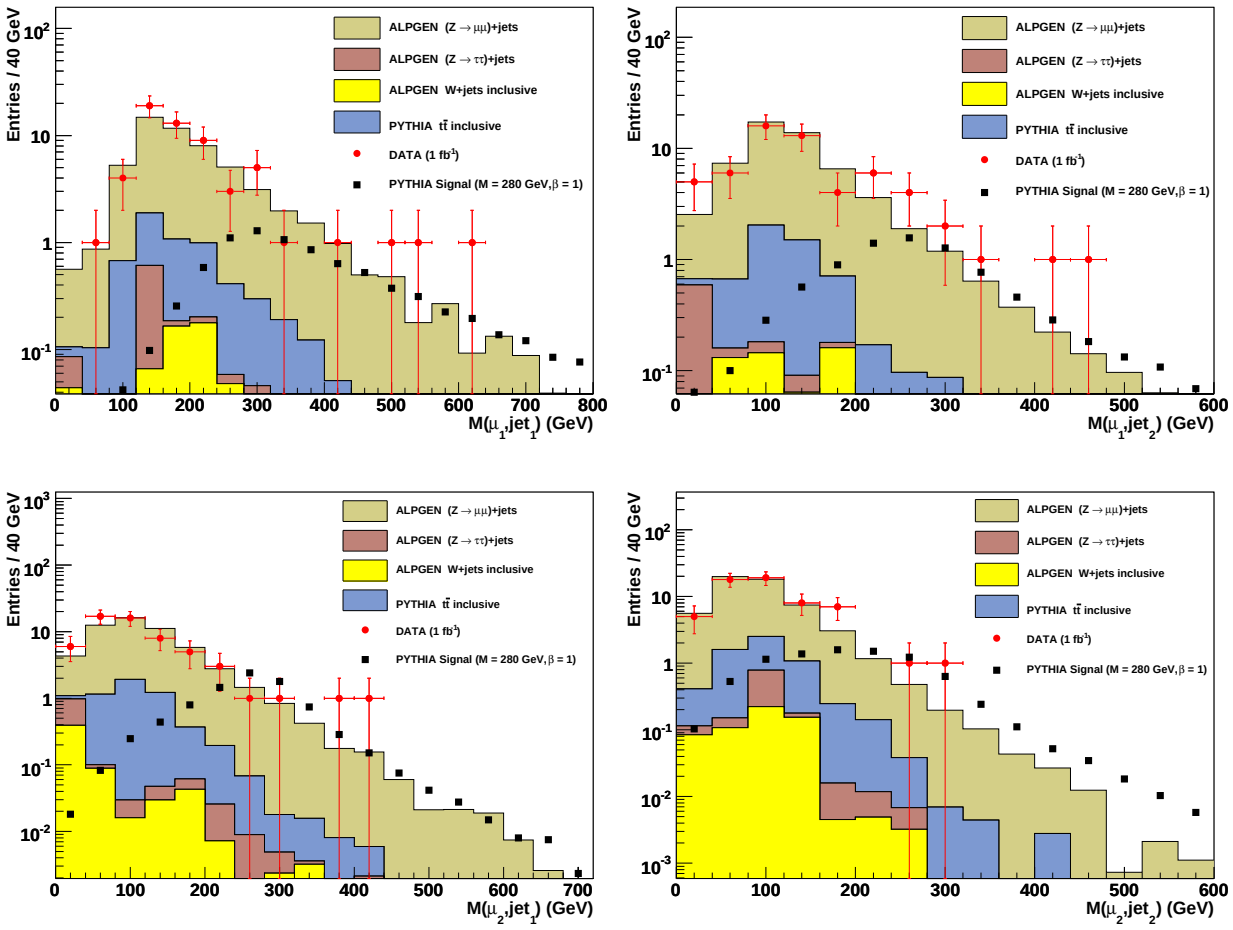

Figure 5.12: *Muon − jet invariant masses constructed from the four vectors of the two leading muons and the two leading jets, after the application of the tight preselection cuts. These variables are contributing to the reconstruction of the leptoquark mass. The distributions of data (red dots), background (histograms), and signal (black squares) are represented. For the signal distribution,* $M_{LQ} = 280\ GeV$ *and* $\beta = 1$ *are assumed.*

expected leptoquark pair production upper cross section limit as function of the lower threshold of a cut placed on the *kNN* output variable. Table 6.3, in the next chapter, provides the limit calculation results of the cut-based analysis. For an assumed leptoquark mass of 280 GeV, an expected limit of 0.018 pb is found, which is to be compared with the lowest expected limit obtained by cutting on the *kNN* output variable, namely 0.013 pb. For the same reasons as described in Section 5.1.3, we divided the *kNN* output variable into 6 bins to further improve the multivariate analysis. The same binning as in the previous case has been used for all the assumed leptoquark masses and considered β values.

The edges and content of each *kNN* bin, including systematic errors, are detailed in Table 5.12, assuming $M_{LQ} = 280\,\text{GeV}$ and $\beta = 1$. The signal efficiencies in each bins, for all the assumed leptoquark masses, and in the case where $\beta = 1$, are provided in Table 5.13. Results relative to the limit calculations based on bins of the *kNN* output variable are discussed in Section 6.3.2).

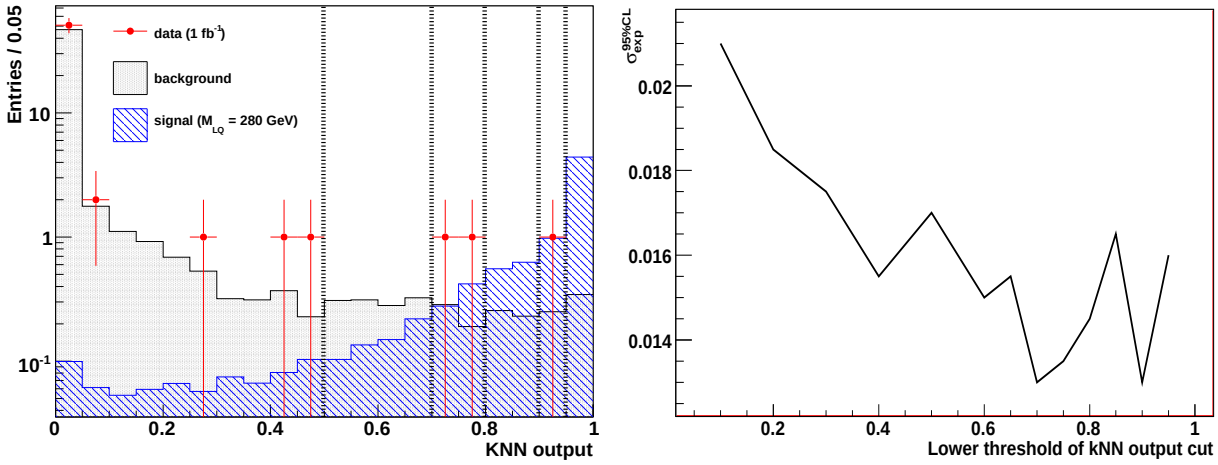

Figure 5.13: *Left: distribution of the kNN output variable for data (red dots), background (black histogram), and signal (blue histogram) events. The chosen bins are separated by the dashed vertical black lines. Right: Expected (95% CL) upper cross section limit on the leptoquark pair production as function of the lower threshold of the one-dimensional cut on the kNN output variable.*

Samples	Bin 1: $0 < kNN < 0.5$	Bin 2: $0.5 < kNN < 0.7$	Bin 3: $0.7 < kNN < 0.8$
$W(+jets) \to l\nu + jets$	$0.57 \pm 0.12 \pm 0.07$	$0.0045 \pm 0.0045 \pm 0.0008$	-
$Z/\gamma^*(+jets) \to \mu\mu + jets$	$48 \pm 1 \pm 6$	$0.99 \pm 0.09 \pm 0.18$	$0.42 \pm 0.04 \pm 0.03$
$Z/\gamma^*(+jets) \to \tau\tau + jets$	$0.66 \pm 0.52 \pm 0.08$	$0.0082 \pm 0.0026 \pm 0.0015$	$0.0029 \pm 0.0011 \pm 0.0002$
$t\bar{t}$ (inclusive)	$4.4 \pm 0.1 \pm 0.8$	$0.22 \pm 0.03 \pm 0.08$	$0.058 \pm 0.018 \pm 0.02$
Total Background	$53 \pm 1 \pm 6$	$1.2 \pm 0.1 \pm 0.2$	$0.48 \pm 0.05 \pm 0.04$
Data	56	0	2
$\epsilon^{\mu\mu qq}_{Signal}$ (%)	3.06	2.57	2.95
N_{signal}	$0.72 \pm 0.02 \pm 0.05$	$0.61 \pm 0.02 \pm 0.02$	$0.70 \pm 0.02 \pm 0.03$

Samples	Bin 4: $0.8 < kNN < 0.9$	Bin 5: $0.9 < kNN < 0.95$	Bin 6: $0.95 < kNN < 1$
$W(+jets) \to l\nu + jets$	$0.0024 \pm 0.0024 \pm 0.0005$	-	-
$Z/\gamma^*(+jets) \to \mu\mu + jets$	$0.42 \pm 0.03 \pm 0.15$	$0.22 \pm 0.03 \pm 0.02$	$0.31 \pm 0.03 \pm 0.04$
$Z/\gamma^*(+jets) \to \tau\tau + jets$	$0.0037 \pm 0.0013 \pm 0.0014$	$0.00039 \pm 0.00028 \pm 0.00004$	$0.0023 \pm 0.0014 \pm 0.0003$
$t\bar{t}$ (inclusive)	$0.058 \pm 0.016 \pm 0.014$	$0.031 \pm 0.010 \pm 0.009$	$0.035 \pm 0.011 \pm 0.007$
Total Background	$0.49 \pm 0.04 \pm 0.16$	$0.25 \pm 0.03 \pm 0.03$	$0.34 \pm 0.03 \pm 0.05$
Data	0	1	0
$\epsilon^{\mu\mu qq}_{Signal}$ (%)	5.01	4.14	18.7
N_{signal}	$1.18 \pm 0.02 \pm 0.04$	$0.98 \pm 0.02 \pm 0.03$	$4.41 \pm 0.05 \pm 0.22$

Table 5.12: *Content of each bin of the kNN variable after the tight preselection. The signal efficiencies (cumulated) for leptoquark decays into the μμqq state are given, as well as the number of events in the data and background samples. The assumed leptoquark mass is 280 GeV, and β is taken equal to 1. First uncertainties are statistical, second are systematical. The systematic uncertainty on the integrated luminosity is not included. "-" refers to the absence of events.*

	$\epsilon^{\mu\mu qq}_{Signal}$ (%)					
M_{LQ} (GeV)	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
140	9.30	3.76	2.30	2.52	1.02	0.311
160	8.92	4.47	2.65	3.60	2.20	1.39
180	8.07	4.79	3.03	4.32	3.35	3.71
200	6.86	5.03	3.36	4.87	4.05	6.28
220	5.85	4.62	3.63	5.10	3.95	9.09
240	4.72	3.96	3.77	5.33	4.20	12.0
260	3.49	3.10	3.35	5.32	4.20	15.9
280	3.06	2.57	2.95	5.01	4.14	18.7
300	2.69	2.34	2.61	4.93	3.91	21.2
320	2.65	2.29	2.64	4.70	3.85	21.8

Table 5.13: *Signal efficiencies (in % and cumulative) after all the cuts, for leptoquark decay into the $\mu\mu jj$ final state, and for all the considered assumed leptoquark masses. The branching fraction β is taken equal to 1.*

5.2.4 Systematic uncertainties

The same treatment as for the $\mu \not{E}_T jj$ channel has been applied for the estimation of the systematic uncertainties. More details and comments on the estimations are provided in Section 5.1.4. The errors relative to the cut-based analysis are given, and for the multivariate analysis, we focused on the case where $M_{LQ} = 280\,\text{GeV}$ and $\beta = 1$.

In the following, only the uncertainties which differ from those of the $\mu \not{E}_T jj$ channel are described.

Muon smearing

For the cut-based analysis, after all selection cuts, we derived relative uncertainties on the muon p_T oversmearing of 9.2%, 10.9% and 0.8% for Z/W production, $t\bar{t}$ background, and signal, respectively.

In the case of the multivariate analysis, the relative uncertainty on the muon p_T oversmearing varies from 0.4% to 4.5% for the signal sample. For the Z/W backgrounds, this error can be as low as 1.5% and take values up to 19%. For the $t\bar{t}$ production, it varies from 0.20% to 25%. Table 5.14 provides the uncertainties relative to the multivariate analysis. Anti-correlations between bins have been taken into account, and are represented by a difference of sign.

Sample	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Z/W	8.7	14	-1.5	19	3.4	11
$t\bar{t}$	-0.20	25	-22	15	14	7.8
Signal	4.5	0.10	3.08	-0.60	-0.44	-3.1

Table 5.14: *Muon transverse momentum smearing uncertainty (in %) on background and signal, in the case where $M_{LQ} = 280\,GeV$ and $\beta = 1$. Anti-correlations between bins are represented by a difference in sign.*

For other (M_{LQ},β) analysis points, uncertainties up to 26% are arising for the Z/W contribution.

The jet shifting, smearing and identification

After the selection of the cut-based analysis, the relative error on the jet shifting is found to be of 6.1%, 2.4%, and 1.0%, for the Z/W backgrounds, the $t\bar{t}$ production, and the signal, respectively. The relative uncertainty on the jet smearing is lower than 0.1% for each simulated sample, and the jet identification error varies from 0.25% to 0.50%.

In the multivariate analysis, relative errors on the jet shifting (reported in Figure 5.7) are found to vary from 0.34% to 11%, from no deviation to 24%, and from 0.34% to 4.7%, for the W/Z production, the $t\bar{t}$ background, and the signal prediction, respectively. In all the background samples, the jet smearing and jet identification induce relative uncertainties below 1% and 2%, respectively, which includes the case where there is no deviation. For the signal prediction, the relative errors due to the jet smearing and jet identification are found to vary from negligible values to 0.27% and 0.65%, respectively.

As well as in the $\mu \not{E}_T jj$ channel, higher values can arise in other (M_{LQ}, β) analysis points, due to statistical fluctuations.

Sample	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
W/Z	7.8	9.3	-0.34	11	-0.75	7.7
$t\bar{t}$	1.9	17	-24	5.6	18	-
Signal	-4.7	-2.3	-0.74	-0.34	0.69	2.4

Table 5.15: *Jet shifting uncertainty (in %) on background and signal, in the case of the multivariate analysis ($M_{LQ} = 280\ GeV$ and $\beta = 1$), and as function of the kNN bins. Anti-correlations between bins are represented by a difference in sign. "-" refers to the absence of deviation.*

The normalization of Z/W samples

An error of 0.5% on the Z and W backgrounds is added to account for uncertainties on their simultaneous normalization to data. This uncertainty has been estimated by varying the normalization window from]60, 120[GeV to]80, 100[GeV.

The ALPGEN jet p_T shape in the Z sample

Similarly to the systematic uncertainty included for the W background in the $\mu \not{E}_T jj$ channel, an additional systematic error is evaluated to account for uncertainties in the correct modeling of the jet p_T shape in Z events. The same method is used.

A mixture composed with up to 43% of PYTHIA has been found to be acceptable. The Figure 5.14 provides the distribution of the second leading jet p_T for ALPGEN, PYTHIA and data, as well as the Kolmogorov-Smirnov test behavior.

In the cut-based analysis, an error of 3.0% is derived for the jet p_T uncertainty. However, the effects of this error when cutting on S_T is and $M(\mu,\mu)^{min}$ are anti-correlated, which lies in the fact that $M(\mu,\mu)^{min}$ includes a correction based on $\not{E}_T$ which depend on the opposite of the jets

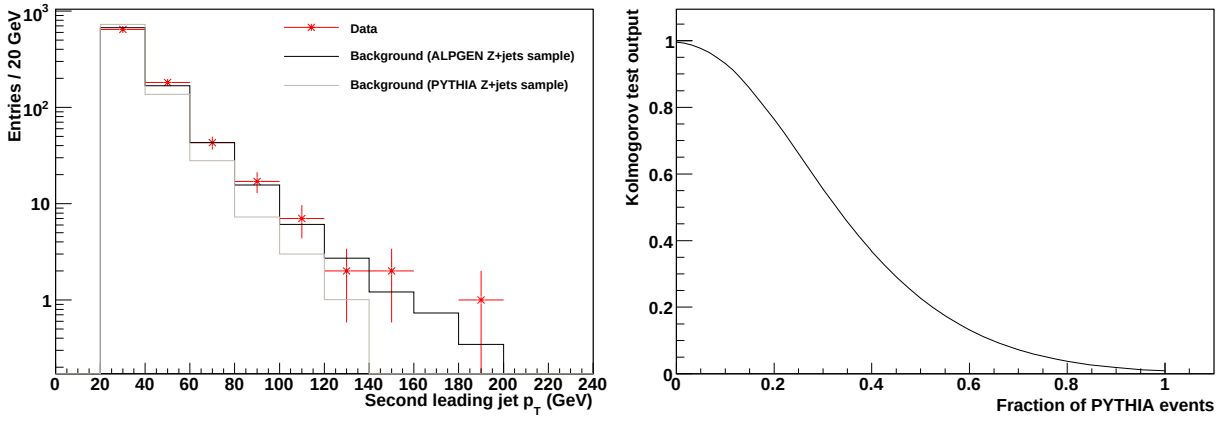

Figure 5.14: *Left: distribution of the second leading transverse momentum in Data and background events, when simulating the Z boson production with* ALPGEN *(black curve) or* PYTHIA *(gray curve).Right: Kolmogorov test output describing the compatibility between the Data and background second leading transverse momentum distribution, as function of the fraction* PYTHIA *events included in addition to* ALPGEN *events for the description of the Z boson production.*

p_T sum. To be conservative, we use the same relative error as for the $\mu \not{E}_T jj$ channel, namely 15%.

The anti-correlation has less influence in the multivariate analysis, since the analysis is divided in multiple *kNN* bins. Relative uncertainties from 3.8% to 29% are derived. The largest calculated errors, among all the (M_{LQ}, β) analysis points, is found to be 41%. The uncertainties due to the ALPGEN jet p_T shape in the case where $M_{LQ} = 280\,\text{GeV}$ and $\beta = 1$ are given in Table 5.16.

Sample	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
jet p_T shape uncertainty	3.8	-7.2	-7.7	-29	9	4.3

Table 5.16: *Uncertainty (in %) due to the* ALPGEN *jet p_T shape in the case of the multivariate analysis, as function of the kNN bins, and assuming $M_{LQ} = 280\,GeV$ and $\beta = 1$. Anti-correlations between bins are represented by a difference in sign.*

Chapter 6

Upper limit on leptoquark pair production cross section and interpretation

Since no excess in data over predicted Standard Model background has been observed, 95% confidence level (CL) limits on the leptoquark pair production cross section have been derived. A semi-Bayesian cross section limit setting method has been used to determine the expected and observed upper cross section bounds on the pair production of scalar leptoquarks decaying either into the final state composed of a muon, a neutrino and two jets, or into the final state composed of two muons and two jets. Limits obtained in the two channels as function of the assumed leptoquark mass and $\beta = Br(LQ \to \mu q)$ are combined and interpreted as lower limits on the leptoquark mass as function of β.

6.1 Limit calculation method

The CL_s method [71], as implemented in the ROOT [72] class *TLimit*, has been utilized for all the upper cross section limits calculated in this thesis.

In this method, the expected number of signal and background events are modeled by Poisson statistics. Two hypotheses are considered: the cases "b" and "s+b", which assume the presence of background only, and background plus signal, respectively. The corresponding confidence levels are referred to as CL_b and CL_{s+b}. They express the probabilities that the observed data events are fully or partially described by background processes only, in the case "b", and signal plus background processes, in the case "s+b". CL_b can thus quantify the confidence of a potential discovery while CL_{s+b} is considered when excluding the possibility of the simultaneous presence of background and new physics.

CL_b and CL_{s+b} are computed by using the likelihood ratio of the two hypotheses for the test statistics. According to the Neyman-Pearson theorem [73], this choice is optimal in the case of completely specified hypotheses and given that the probability densities of both hypotheses are identical, which is the case here. The likelihood ratio, for a single counting experiment, can be written as follows:

$$\mathcal{L}_{ratio} = \frac{e^{-(s+b)}\,(s+b)^d/d!}{e^{-b}\,b^d/d!} \tag{6.1}$$

where $\mathcal{L}_{ratio}$ is the likelihood ratio, and s, b and d, the number of signal, background and data events, respectively.

Under their respective hypotheses, the confidence levels CL_{s+b} and CL_b are derived by computing the probability that $\mathcal{L}_{ratio}$ is lower or equal than that observed in the data.

When the background fluctuates to low values, very small CL_{s+b} and CL_b are calculated. Thus it is not possible to conclude that the signal has a low rate or that it can be excluded. To account for this effect, the quantity CL_s, defined as the ratio of both confidence levels, is expressed as follows:

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{6.2}$$

The observed 95% confidence level upper bound on the leptoquark pair production cross section is obtained when the respective CL_s value is equal to 5%. By assuming the presence of background only, expected limits are computed, which provide a discriminant quantifying how well the analysis can exclude the presence of signal, given that it is not observed. For the calculation of expected limits, pseudo-data, which consist of events distributed according to a Poisson centered in b, are used.

By using $\mathcal{L}_{ratio}$ as the test statistics, the statistic for the outcome of n channels is then simply the product of the n test statistics of the channels taken separately. Thus the combination of the 6 *kNN* bins of each (M_{LQ},β) analysis case, when all considered as different channels, can be easily implemented with this method. As well, the combination of the $\mu\not{E}_T jj$ and $\mu\mu jj$ channels for a given (M_{LQ},β) analysis point can be achieved by considering the test statistics of the corresponding 12 *kNN* bins.

Systematic uncertainties have been taken into account in the limit calculation, including their correlations between each bin (or channel). Because of the implementation of the CL_s method, all the systematic errors have been assumed to be Gaussian in Sections 5.1.4 and 5.2.4.

In the following sections, the expected and observed upper cross section limits have been calculated given the remaining number of signal and background events after all cuts, and their respective uncertainties. The upper cross section bounds are compared to the NLO prediction of the leptoquark pair production cross section, reduced by its uncertainty [74]. The leptoquark mass for which the upper cross section limit is equal to the lower bound of the theoretical prediction can be interpreted as the lower limit on the leptoquark mass. The error on the cross section includes the factorization and renormalization scale uncertainty (varied between $\frac{1}{2}M_{LQ}$ and $2M_{LQ}$), reduced by the use of NLO theoretical predictions, and the PDF uncertainty (evaluated using the CTEQ6.1M error PDF sets). Both uncertainties have been added in quadrature, for each assumed leptoquark mass.

The best lower limit for the leptoquark mass is thus obtained when minimizing the upper bound on the leptoquark pair production cross section. The *kNN* bins with the highest signal over background ratio have the largest contribution to the limit minimization. The fact of dividing the *kNN* output distribution into 6 bins allows to use shape information in addition, and thus provide "channels" with better signal over background ratios in the signal-like region.

6.2 Limits in the $\mu \not{E}_T jj$ channel

Both $\mu\nu qq$ and $\mu\mu qq$ generated events have been included when deriving the signal rate, which is therefore a mixture of events weighted with the branching ratios of the $\mu \not{E}_T jj$ and $\mu\mu jj$ channels. As a result, no limit on the leptoquark pair production cross section times branching ratio can be evaluated. The upper cross section limit can only be given with respect to the leptoquark pair production itself (i.e. without branching fractions), and has to be compared to the theoretical pair production cross section.

The practical drawback lies in the fact that the upper cross section limits for various values of β cannot be derived by simply scaling the theoretical cross section with the branching ratio and comparing it to the estimated upper cross section limit. Therefore, for each considered β value, the limit setting procedure has to be performed separately, as achieved in the case of the multivariate analysis.

6.2.1 Limits for the cut-based analysis

Observed and expected lower limits on the leptoquark mass of both 213 GeV have been estimated, assuming $\beta = 0.5$. Thus, the cut-based analysis improves over previous mass limits for second generation scalar leptoquarks. Previous limits obtained in the $\mu\nu jj$ channel at $\beta = 0.5$ are $M_{LQ} > 170$ GeV (CDF, Run II [38]) and $M_{LQ} > 160$ GeV (DØ Run I [75]). The comparison between the theoretical leptoquark pair production cross section (listed in Table 3.2) and the expected and observed upper cross section limits is represented in Figure 6.1. The associated Table 6.1 provides the upper cross section limits values for each assumed leptoquark mass. The fact the observed and expected limits are identical can be explained by the good agreement between the data (8 events) and the Standard Model background (8.6 events).

6.2.2 Limits for the multivariate analysis

In the case of the $\mu \not{E}_T jj$ channel, the branching ratio for $\mu\nu qq$ final states vanishes for $\beta = 0$ and $\beta = 1$. However for $\beta = 1$, $\mu\mu qq$ final states still contribute to the signal rate as one of the two muons could be non-identified. Limits have therefore been calculated for $0.1 \leq \beta \leq 1$ (with a step of 0.1). Table 6.2 lists all the derived expected and observed 95% CL upper limits on the leptoquark pair production cross section. For $\beta = 0.5$, we found an observed (expected) lower bound on the leptoquark mass of 248 GeV (242 GeV). Results are thus significantly improved compared to those of the cut-based analysis. Figure 6.2 shows the comparison between the upper cross section limits and the theoretical prediction, assuming $\beta = 0.5$.

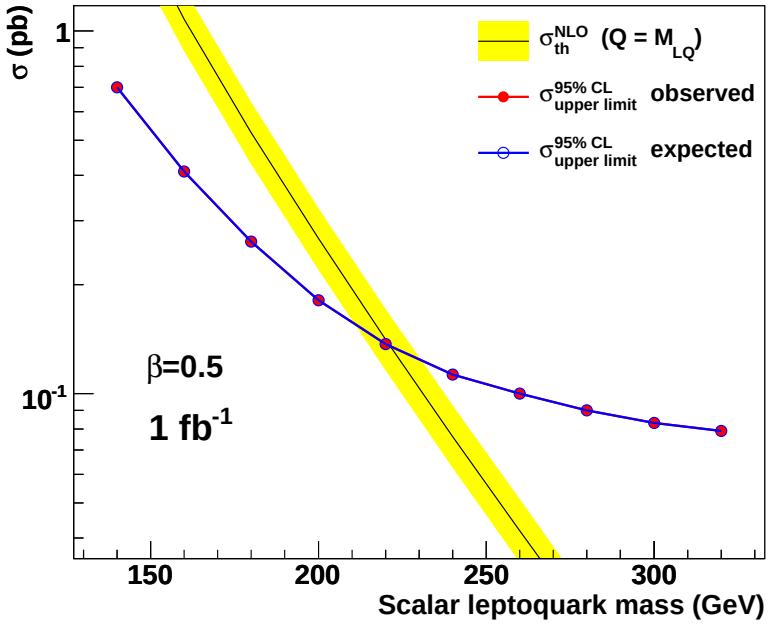

Figure 6.1: *Cut-based analysis, $\mu \not{E}_T jj$ channel. Comparison between the 95% CL upper limit on the leptoquark pair production cross section and the theoretical pair production cross section (NLO). By considering the lower bound of the NLO prediction, the mass of a second generation leptoquark can be ruled out up to 213 GeV. The expected mass limit is also 213 GeV. Bounds are derived assuming $\beta = 0.5$.*

	$\beta = 0.5$	
M_{LQ} (GeV)	$\sigma^{expected}_{95\%\,CL}$ (pb)	$\sigma^{observed}_{95\%\,CL}$ (pb)
140	0.70	0.70
160	0.41	0.41
180	0.26	0.26
200	0.18	0.18
220	0.14	0.14
240	0.11	0.11
260	0.10	0.10
280	0.090	0.090
300	0.083	0.083
320	0.079	0.079

Table 6.1: *Cut-based analysis, $\mu \not{E}_T jj$ channel. Expected and observed 95% CL upper limits on the leptoquark pair production cross section, as function of the assumed leptoquark mass, and for $\beta = 0.5$.*

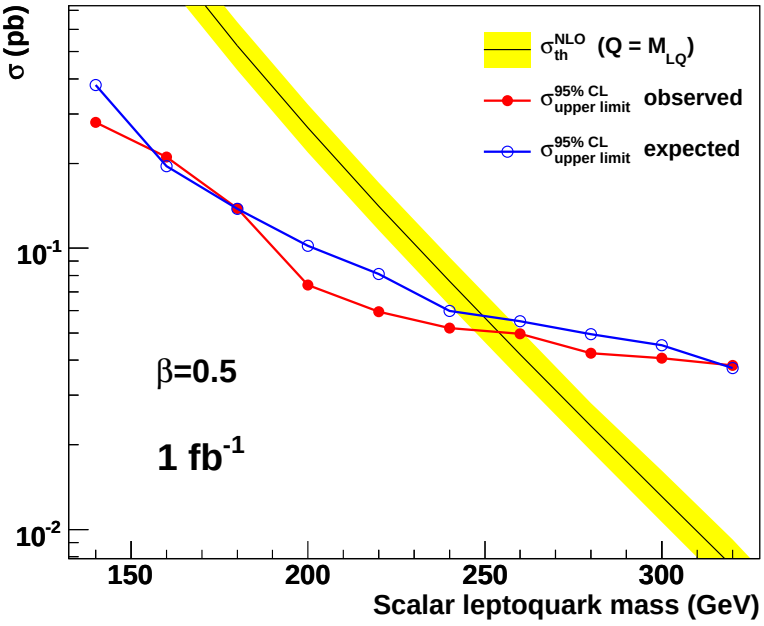

Figure 6.2: *Multivariate analysis, $\mu\not{E}_T jj$ channel. Comparison between the 95% CL upper limit on the leptoquark pair production cross section and the theoretical pair production cross section (NLO). By considering the lower errors of the NLO prediction, the mass of a second generation leptoquark can be ruled out up to 248 GeV. The expected mass limit is 242 GeV. Limits are derived assuming $\beta = 0.5$.*

	$\beta = 0.5$	
$M_{LQ}(GeV)$	$\sigma^{expected}_{95\%\,CL}$ (pb)	$\sigma^{observed}_{95\%\,CL}$ (pb)
140	0.38	0.28
160	0.20	0.21
180	0.14	0.14
200	0.10	0.074
220	0.081	0.060
240	0.060	0.052
260	0.055	0.050
280	0.049	0.042
300	0.045	0.041
320	0.038	0.038

Table 6.2: *Multivariate analysis, $\mu\not{E}_T jj$ channel. Expected and observed 95% CL upper limits on the leptoquark pair production cross section, as function of the assumed leptoquark mass, and for $\beta = 0.5$.*

6.3 Limits in the $\mu\mu jj$ channel

In contrast to the $\mu \not{E}_T jj$ channel, only $\mu\mu qq$ generated events have been considered in this case, as the muon fake rate proved to be negligible under the constraints of the preselection. Therefore, all the signal events have been weighted by the same branching ratio, namely β^2. It is thus possible to estimate an upper limit both on the leptoquark pair production cross section and on the leptoquark pair production cross section times branching ratio. The latter alternative is preferred in this section in order to simplify the estimation of the mass limits for all the β values.

As a result we calculated the 95% CL upper limit on the leptoquark pair production cross section times branching ratio only once. Then, we derived the leptoquark mass limits for all the β values by simply comparing the estimated upper cross section limits with the theoretical leptoquark pair production cross section scaled by β^2.

6.3.1 Limits for the cut-based analysis

We derived an observed (expected) lower limit of 290 GeV (283 GeV) on the leptoquark mass when assuming $\beta = 1$, and 234 GeV (224 GeV) if $\beta = 0.5$. These measurements are significantly improved in comparison with the previous mass limits in this channel. Previous limits from the DØ experiment (Run II), derived with an integrated luminosity of 294 pb^{-1}, are $M_{LQ} > 247$ GeV and $M_{LQ} > 182$ GeV, for $\beta = 1$ and $\beta = 0.5$, respectively [35]. At the CDF Experiment (Run II), the latest results for this channel were obtained with an integrated luminosity of 198 pb^{-1}, and excluded leptoquarks with a mass below 224 GeV [38]. Expected and observed upper limits on the leptoquark pair production cross section times branching ratio are provided in Table 6.3. Figure 6.3 shows the comparison between the upper cross section limits and the theoretical prediction times branching fraction, represented for $\beta = 1$ and $\beta = 0.5$.

6.3.2 Limits for the multivariate analysis

As for the $\mu \not{E}_T jj$ channel, limits have been determined for $0.1 \leq \beta \leq 1$ (with a step of 0.1). Observed (expected) lower thresholds on the leptoquark mass of 299 GeV (302 GeV) and 234 GeV (244 GeV) have been obtained when assuming $\beta = 1$ and $\beta = 0.5$, respectively. The upper cross section limits are improved over the cut-based analysis, while not providing a gain comparable to what is observed for the $\mu \not{E}_T jj$ channel. This could be explained by the fact that the cut on $M(\mu, \mu)^{min}$ already well discriminates the signal from the background. Table 6.4 lists the derived upper cross section bounds, and Figure 6.4 compares them to the theoretical prediction assuming $\beta = 0.5$ or $\beta = 1$.

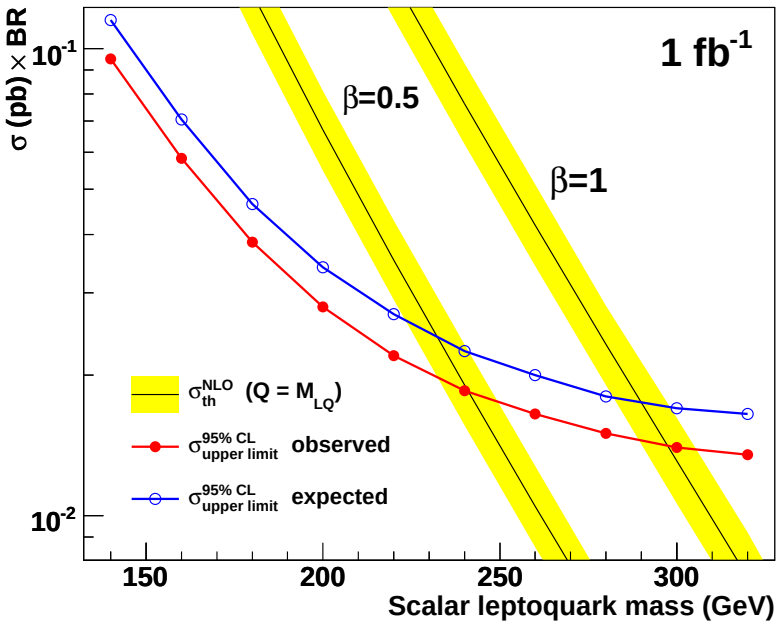

Figure 6.3: *Cut-based analysis, $\mu\mu jj$ channel. Comparison between the 95% CL upper limit on the leptoquark pair production cross section times branching ratio and the theoretical pair production cross section (NLO) times branching ratio. By considering the lower errors of the NLO prediction times branching ratio, and if assuming $\beta = 1$, the mass of a second generation leptoquark can be ruled out up to 290 GeV. The expected mass limit is 283 GeV. In the case $\beta = 0.5$, the observed (expected) mass limit is 234 GeV (224 GeV).*

M_{LQ} (GeV)	$\sigma_{95\%}^{expected}$ (pb)	$\sigma_{95\%}^{observed}$ (pb)
140	0.115	0.095
160	0.071	0.058
180	0.047	0.039
200	0.034	0.028
220	0.027	0.022
240	0.023	0.019
260	0.020	0.017
280	0.018	0.015
300	0.017	0.014
320	0.017	0.014

Table 6.3: *Cut-based analysis, $\mu\mu jj$ channel. Expected and observed 95% CL upper limits on the leptoquark pair production cross section times branching ratio, depending on the assumed leptoquark mass.*

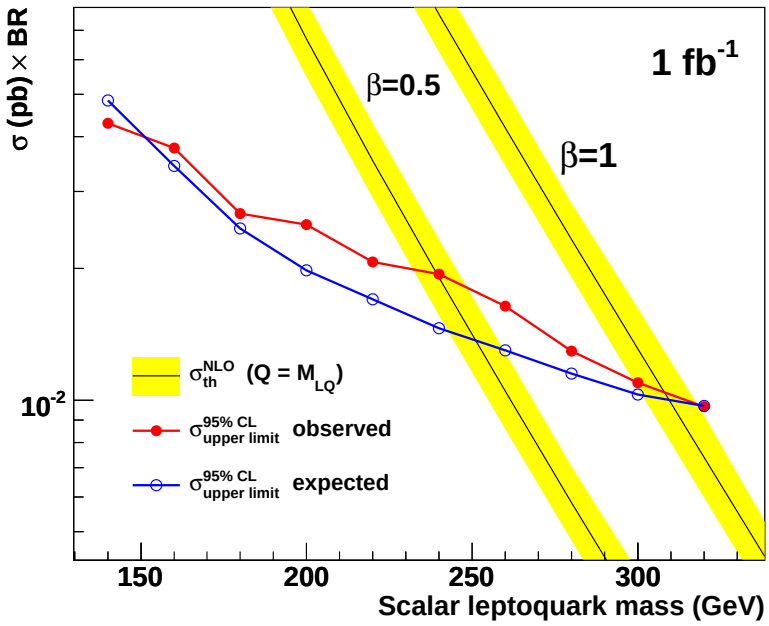

Figure 6.4: *Multivariate analysis, $\mu\mu jj$ channel. Comparison between the 95% CL upper limit on the leptoquark pair production cross section times branching ratio and the theoretical pair production cross section (NLO) times branching ratio. By considering the lower uncertainties of the NLO prediction, and if assuming $\beta = 1$, the mass of a second generation leptoquark can be ruled out up to 299 GeV. The expected mass limit is 302 GeV. In the case $\beta = 0.5$, the observed (expected) mass limit is 234 GeV (244 GeV).*

M_{LQ} (GeV)	$\sigma_{95\%}^{expected}$ (pb)	$\sigma_{95\%}^{observed}$ (pb)
140	0.048	0.043
160	0.034	0.038
180	0.025	0.027
200	0.020	0.025
220	0.017	0.021
240	0.015	0.019
260	0.013	0.016
280	0.012	0.013
300	0.010	0.011
320	0.010	0.010

Table 6.4: *Multivariate analysis, $\mu\mu jj$ channel. Expected and observed 95% CL upper limits on the leptoquark pair production cross section times branching ratio, as function of the assumed leptoquark mass.*

6.4 Combination of $\mu\not{E}_T jj$ and $\mu\mu jj$ channels

The upper limits on the leptoquark pair production cross section obtained from the combination of both channels are evaluated as function of β, varied between 0.1 to 1 (with a step of 0.1), and the assumed leptoquark mass, taking values between 140 and 320 GeV (with a step of 20 GeV). As in Section 6.2, the calculated upper cross section limits are not bounds on the leptoquark pair production cross section times branching fraction, but limits on the leptoquark pair production cross section itself.

Assuming $\beta = 0.5$ and $\beta = 1$, leptoquarks decaying to a muon and a quark can be ruled out up to 262 GeV and 309 GeV, respectively, as shown in Figure 6.5. Expected lower mass limits are 310 GeV and 267 GeV, respectively. Results of both channels taken separately are thus improved, as summarized in Table 6.5. Tables 6.6 and 6.7 provide the comparison between the upper bounds on the leptoquark pair production cross section determined with the different event selections considered, for $\beta = 0.5$ and $\beta = 1$, respectively. In the case $\beta = 0.1$, the contribution of the $\mu\mu jj$ channel turns out to be negligible, and the lower limit on the leptoquark mass corresponds to the one determined for the $\mu\not{E}_T jj$ channel, namely 174 GeV.

The branching fraction of the $\mu\mu qq$ state is maximized for $\beta = 1$, while the $\mu\nu qq$ production is maximized for $\beta = 0.5$. The case $\beta = 0$ can be investigated by studying the $\nu\nu jj$ channel [33], which goes beyond the scope of this thesis. As explained in sections 5.1.3 and 5.2.3, the optimization of the event selection is done for each point of the (β, M_{LQ}) space analyzed. As well, limits on the leptoquark pair production cross section, interpreted as bounds on the leptoquark mass, are derived for each β value analyzed. Figure 6.6 provides the exclusion plot for second generation scalar leptoquarks, as function of β and M_{LQ}. Table 6.8 lists the corresponding lower mass limits.

Figure 6.6 shows remarkable consequences of the mixture of signal events considered in the $\mu\not{E}_T jj$ channel. First, the limit on the leptoquark mass derived in the $\mu\not{E}_T jj$ channel is not symmetric with respect to $\beta = 0.5$ as it is expected when only the $\mu\nu qq$ state, of branching fraction $2\beta(1-\beta)$, is included. Furthermore, for $\beta = 1$, the $\mu\not{E}_T jj$ channel contributes to the leptoquark mass limit, since we included $\mu\mu qq$ events (with one muon not reconstructed). In addition, the best sensitivity for the $\mu\not{E}_T jj$ channel is obtained for $\beta = 0.6$, although $\beta = 0.5$ maximizes the branching ration of the $\mu\nu qq$ state. The inclusion of $\mu\mu qq$ events in the $\mu\not{E}_T jj$ channel thus proved to sensibly improve the analysis sensitivity.

	$\mu\nu jj$ channel		$\mu\mu jj$ channel		Combination	
β	$M_{LQ}^{expected}$	$M_{LQ}^{observed}$	$M_{LQ}^{expected}$	$M_{LQ}^{observed}$	$M_{LQ}^{expected}$	$M_{LQ}^{observed}$
0.1	175	174	-	-	175	174
0.5	242	248	244	234	267	262
1	203	224	302	299	310	309

Table 6.5: *Expected and observed lower bounds (95% CL) on the second generation leptoquark mass (in GeV) depending on the channel. The cases $\beta = 1$, $\beta = 0.5$ and $\beta = 0.1$ are shown. For $\beta = 0.1$, the contribution of the $\mu\mu jj$ channel is negligible, and no limit has been calculated, which is denoted by "-".*

$\mu \not\!\!E_T jj$ channel			Combination
Cut-based	Cut on kNN	Multivariate (6 bins)	Multivariate (12 bins)
0.18	0.13	0.10	0.053

Table 6.6: *Expected upper limits (95% CL) on the second generation leptoquark pair production cross section (in pb) assuming $\beta = 0.5$ and $M_{LQ} = 200\,\text{GeV}$, for all the event selections considered. Results for the $\mu \not\!\!E_T jj$ channel and for the combination of both channels are shown.*

$\mu\mu jj$ channel			Combination
Cut-based	Cut on kNN	Multivariate (6 bins)	Multivariate (12 bins)
0.018	0.013	0.012	0.0089

Table 6.7: *Expected upper limits (95% CL) on the second generation leptoquark pair production cross section times branching ratio (in pb) assuming $M_{LQ} = 280\,GeV$, for all the event selections considered. Results for the $\mu\mu jj$ channel and for the combination of both channels are shown in the case $\beta = 1$.*

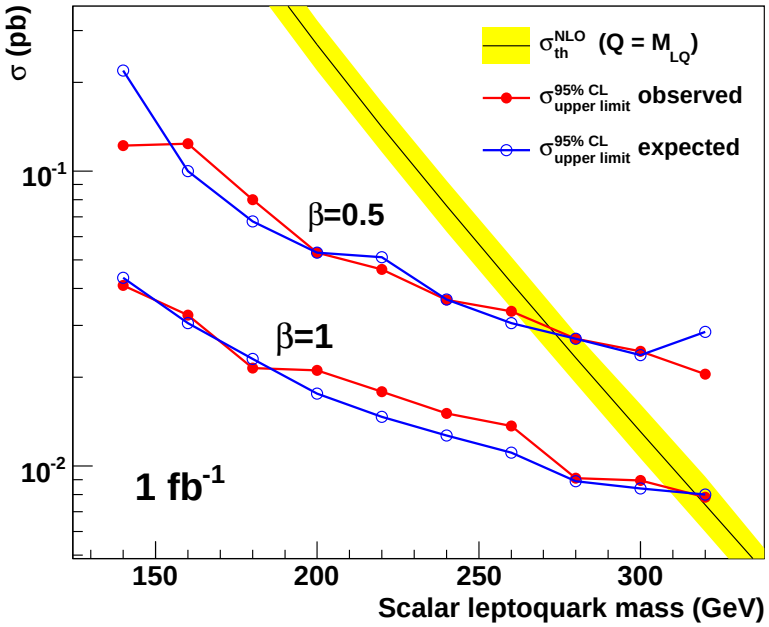

Figure 6.5: *Multivariate analysis, combination of the $\mu\nu jj$ and $\mu\mu jj$ channels. 95% CL upper limits on the leptoquark pair production cross section are compared to the theoretical pair production cross section (NLO). By considering the lower bound of the NLO prediction, we derive an observed (expected) lower limit on the mass of second generation leptoquarks of 309 GeV (310 GeV) and 262 GeV (267 GeV), when assuming $\beta = 1$ and $\beta = 0.5$, respectively.*

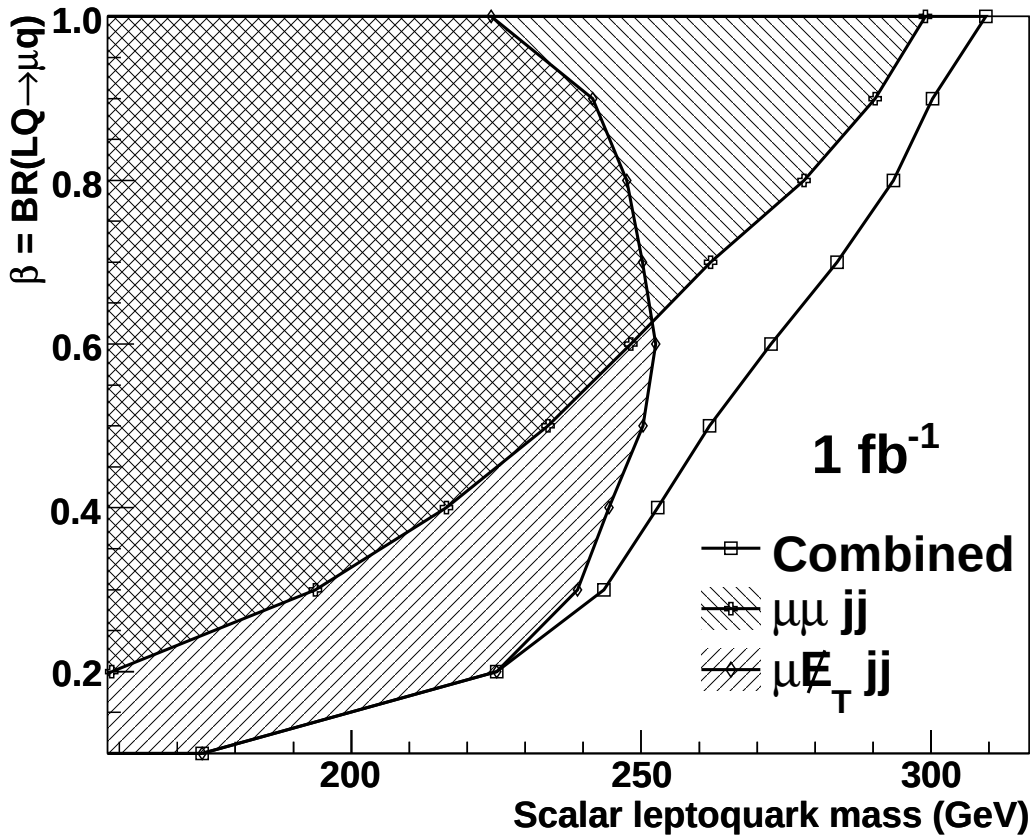

Figure 6.6: *Multivariate analysis, combination of the $\mu\nu jj$ and $\mu\mu jj$ channels. Lower mass limits (in GeV) on second generation leptoquarks are shown as function of $\beta = Br(LQ \to \mu q)$.*

β	$M_{LQ}^{expected}$	$M_{LQ}^{observed}$
0.1	175	174
0.2	211	225
0.3	236	244
0.4	252	253
0.5	266	262
0.6	275	272
0.7	284	284
0.8	295	294
0.9	302	300
1	310	309

Table 6.8: *Multivariate analysis, combination of the $\mu\nu jj$ and $\mu\mu jj$ channels. Observed and expected lower limits on the second generation leptoquark mass (in GeV), as function of β.*

Conclusion

The connection between the lepton and quark sectors of the Standard Model can be explained by extended theories, in which a new type boson, the leptoquark (LQ), is introduced naturally. Leptoquarks carry both lepton and baryon quantum numbers and thus allow lepton-quark transitions. Due to experimental constraints, searches for leptoquarks at hadron collider are commonly carried out assuming an effective model which constrains the leptoquarks to couple only to leptons and quarks of the same fermion generation. As a consequence, three generations of leptoquarks are distinguished, depending on the generation of the fermions to which they couple.

A search for the pair production of second generation scalar leptoquarks has been carried out at the Tevatron $p\overline{p}$ collider, operating at $\sqrt{s} = 1.96\,\text{TeV}$. Using $1\,\text{fb}^{-1}$ of data collected by the DØ detector between August 2002 and February 2006, two channels have been investigated. In one case, one of the leptoquarks is assumed to decay to a muon and a quark, with a branching fraction β, and the other one to a neutrino and a quark, which leads to a final state composed of a muon, missing transverse energy $\not\!\!E_T$ and two jets. In the other case, both leptoquarks are assumed to decay to a muon and a quark, which leads to an event topology consisting of two muons and two jets. Simulated signal events arising from the $LQ\overline{LQ} \to \mu\mu qq$ decay mode have also been considered in the study of the $\mu\not\!\!E_T jj$ channel, given that one muon might not be reconstructed. As the muon fake rate is negligible in the analysis presented, the $LQ\overline{LQ} \to \mu\nu qq$ decay mode has not been studied for the $\mu\mu jj$ channel.

Improved data statistics have been obtained by selecting the data with an inclusive OR of various single muon triggers. A method to perform such a combination in the case of prescaled triggers has been detailed. The $\mu\not\!\!E_T jj$ has been selected with 23 single muons triggers, while for the $\mu\mu jj$ channel, only 8 triggers have been used, due to correlations between the two muons.

In both channels, collected data events have been preselected with quality criteria and compared to the Standard Model prediction. A tighter selection has been imposed to discriminate leptoquark candidate events from the Standard Model background, which mainly consists of W and Z/γ^* boson production for the $\mu\not\!\!E_T jj$ and $\mu\mu jj$ channels, respectively. The event selection has first been achieved by applying cuts on kinematic and reconstructed variables, and has been further improved by the use of the k-Nearest-Neighbors algorithm (k-NN) as a multivariate classifier. The k-NN algorithm is based on the estimation of a signal-like probability within an adaptive volume constructed in as much dimensions as there are physics variables considered. In order to enhance the separation between signal and background, and to account for their relative shapes, the selection based on the multivariate classifier consisted in the division of the k-NN output variable into six bins.

Since no excess of data over the expected Standard Model background has been observed, 95% confidence level limits on the leptoquark pair production cross section have been calculated, as function of the leptoquark mass and β. By comparing these upper bounds to the theoretical

prediction reduced by its uncertainty, lower limits on the leptoquark mass have been derived as function of β.

In the case of the $\mu \not{E}_T jj$ channel alone, the mass of a leptoquark decaying to a muon and a quark or to a neutrino and a quark has been ruled out up to 213 GeV, assuming $\beta = 0.5$. The multivariate analysis improved this lower limit to 248 GeV.

In the case of the $\mu\mu jj$ channel alone, second generation leptoquarks with a mass lower than 290 GeV (234 GeV) have been excluded for $\beta = 1$ ($\beta = 0.5$). The multivariate analysis increased this limit to 299 GeV assuming $\beta = 1$, while no improvement has been observed for $\beta = 0.5$.

The combination of both channels, performed for values of β between 0.1 and 1, improves the lower limit on the second generation leptoquark mass to 309 GeV and 262 GeV for $\beta = 1$ and $\beta = 0.5$, respectively. In the case $\beta = 0.1$, a lower bound on the leptoquark mass of 174 GeV is derived. All the limits obtained for $\beta \geq 0.1$ are the best to date.

Appendix A

Definition of single muon trigger terms

Table A summarizes the definitions of the terms which are part of the single muon triggers combined in chapter 4.

The "all" (A), "wide" (W) and "forward" (B) regions are defined by $|\eta| < 2.0$ (muon system coverage), $|\eta| < 1.5$ (CFT system coverage), and $1.0 \leq |\eta| < 2.0$.

Trigger term	**Criteria**
L1WTN0	level-1 muon tight scintillator, in wide region
L1ATL0	level 1 muon tight scintillator, loose wire, in all region
L1ATN0	level 1 muon tight scintillator, in all region
L1WTL10	level 1 muon loose wire and tight scintillator, $p_T > 10$ GeV, in wide region
L1WTN10	level 1 muon tight scintillator, $p_T > 10$ GeV, in wide region
L1WLL10	level 1 muon loose scintillator and loose wire, $p_T > 10$ GeV, in wide region
L1WTT0	level 1 muon tight scintillator and tight wire, in wide region
L1BTT0	level 1 muon tight scintillator and tight wire, in forward region
L1TRK10	CFT track with $p_T > 10$ GeV
L1ITRK10	CFT isolated track with $p_T > 10$ GeV
L2Mx	muon of level 2 medium quality, $p_T > x$ GeV
L3Lx	muon of level 3 loose quality, and fulfilling $p_T > x$ GeV
L3TRKx	global track in CFT or SMT with $p_T > x$ GeV
L3TRKxH10	global track in CFT or SMT with 10 hits and $p_T > x$ GeV
L3CMLx	level 3 loose muon matched to a central track, with $p_T > x$ GeV
ISO_MUON_CAL3	nearby tracks and calorimeter cells (cone 0.1) energies lower than 3 GeV

Table A.1: *Definitions of single muon trigger terms, at level 1, level 2, and level 3.*

Bibliography

[1] S. Weinberg, Eur. Phys. J. **C34**, 5 (2004), [hep-ph/0401010], and references therein.

[2] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

[3] J. C. Pati and A. Salam, Phys. Rev. **D10**, 275 (1974).

[4] W. Buchmüller and D. Wyler, Phys. Lett. **B177**, 377 (1986).

[5] E. Eichten, K. D. Lane and M. E. Peskin, Phys. Rev. Lett. **50**, 811 (1983).

[6] E. Eichten, I. Hinchliffe, K. D. Lane and C. Quigg, Phys. Rev. **D34**, 1547 (1986).

[7] H1 Collaboration, C. Adloff *et al.*, Z. Phys. **C74**, 191 (1997), [hep-ex/9702012].

[8] ZEUS, J. Breitweg *et al.*, Z. Phys. **C74**, 207 (1997), [hep-ex/9702015].

[9] CDF, C. Grosso-Pilcher, G. L. Landsberg and M. Paterno, hep-ex/9810015.

[10] H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).

[11] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973).

[12] H. Fritzsch, M. Gell-Mann and H. Leutwyler, Phys. Lett. **B47**, 365 (1973).

[13] S. L. Glashow, Nucl. Phys. **22**, 579 (1961).

[14] A. Salam and J. C. Ward, Phys. Lett. **13**, 168 (1964).

[15] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[16] P. W. Higgs, Phys. Lett. **12**, 132 (1964).

[17] H. Georgi and S. L. Glashow, Phys. Rev. **D6**, 429 (1972).

[18] P. Langacker, Phys. Rept. **72**, 185 (1981).

[19] W.-M.Yao et al. (Particle Data Group), J. Phys. G 33, 1 (2006) and 2007 partial update for the 2008 edition.

[20] J. C. Pati and A. Salam, Phys. Rev. **D8**, 1240 (1973).

[21] J. Wess and B. Zumino, Phys. Lett. **B49**, 52 (1974).

[22] S. P. Martin, hep-ph/9709356.

[23] G. Jungman, M. Kamionkowski and K. Griest, Phys. Rept. **267**, 195 (1996), [hep-ph/9506380].

[24] H. K. Dreiner, Pramana **51**, 123 (1998).

[25] M. Kuze and Y. Sirois, Prog. Part. Nucl. Phys. **50**, 1 (2003), [hep-ex/0211048].

[26] W. Buchmüller, R. Rückl and D. Wyler, Phys. Lett. **B191**, 442 (1987).

[27] J. Blümlein and R. Rückl, Phys. Lett. **B304**, 337 (1993).

[28] J. Blümlein, E. Boos and A. Kryukov, Phys. Lett. **B392**, 150 (1997), [hep-ph/9610506].

[29] M. Krämer, T. Plehn, M. Spira and P. M. Zerwas, Phys. Rev. Lett. **79**, 341 (1997), [hep-ph/9704322].

[30] J. Pumplin *et al.*, JHEP **07**, 012 (2002), [hep-ph/0201195].

[31] P. Calfayan, on behalf of the DØ and CDF collaborations, in proceedings of the European Physical Society conference, Manchester (2007).

[32] H1 Collaboration, A. Aktas *et al.*, Phys. Lett. **B629**, 9 (2005), [hep-ex/0506044].

[33] DØ Collaboration, V. M. Abazov *et al.*, Phys. Lett. **B640**, 230 (2006), [hep-ex/0607009].

[34] DØ Collaboration, V. M. Abazov *et al.*, Phys. Lett. **B647**, 74 (2007), [hep-ex/0612012].

[35] DØ Collaboration, V. M. Abazov *et al.*, Phys. Lett. **B636**, 183 (2006), [hep-ex/0601047].

[36] CDF Collaboration, D. E. Acosta *et al.*, Phys. Rev. **D72**, 051107 (2005), [hep-ex/0506074].

[37] DØ Collaboration, V. M. Abazov *et al.*, Phys. Rev. **D71**, 071104 (2005), [hep-ex/0412029].

[38] CDF Collaboration, A. Abulencia *et al.*, Phys. Rev. **D73**, 051102 (2006), [hep-ex/0512055].

[39] DØ Collaboration, Search for third generation scalar leptoquarks in the tau-b-tau-b final state, Conference Note 5370 (2007).

[40] DØ Collaboration, V. M. Abazov *et al.*, Phys. Rev. Lett. **99**, 061801 (2007), [0705.0812].

[41] CDF Collaboration, T. Aaltonen *et al.*, 0706.2832.

[42] J. Thompson, FERMILAB-TM-1909.

[43] G. Jackson, FERMILAB-TM-1991.

[44] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **74**, 2632 (1995), [hep-ex/9503003].

[45] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **74**, 2626 (1995), [hep-ex/9503002].

[46] CDF Collaboration, A. Abulencia *et al.*, Phys. Rev. Lett. **98**, 161801 (2007), [hep-ex/0702027].

[47] CDF Collaboration, T. Aaltonen *et al.*, 0801.4806.

[48] DØ Collaboration, V. M. Abazov *et al.*, Phys. Rev. Lett. **98**, 181802 (2007), [hep-ex/0612052].

[49] DØ Collaboration, V. M. Abazov *et al.*, Nucl. Instrum. Meth. **A565**, 463 (2006), [physics/0507191].

[50] DØ Collaboration, S. Abachi, FERMILAB-PUB-96-357-E.

[51] V. M. Abazov *et al.*, Nucl. Instrum. Meth. **A552**, 372 (2005), [physics/0503151].

[52] S. Agostinelli *et al.*, Nucl. Instrum. Meth. **A506**, 250 (2003).

[53] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, JHEP **07**, 001 (2003), [hep-ph/0206293].

[54] T. Sjostrand *et al.*, Comput. Phys. Commun. **135**, 238 (2001), [hep-ph/0010017].

[55] R. Hamberg, W. L. van Neerven and T. Matsuura, Nucl. Phys. **B359**, 343 (1991).

[56] N. Kidonakis and R. Vogt, Phys. Rev. **D68**, 114014 (2003), [hep-ph/0308222].

[57] G. Blazey *et al.*, in proceedings of the workshop "qcd and weak boson physics in Run II" edited by U. Baur, R.K. Ellis, and D. Zeppenfeld, Batavia, p.47 (2000).

[58] A. Harel, Jet id optimisation, DØ Note 4919 (2005).

[59] Jet Energy Scale Group, Jet energy scale determination at DØ Run II (final p17 version), DØ Note 5382 (2007).

[60] N. Makovec and J.-F. Grivaz, The relative data-monte-carlo jet energy scale, DØ Note 4807 (2005).

[61] N. Makovec and J.-F. Grivaz, Shifting, smearing and removing simulated jets, DØ Note 4914 (2005).

[62] M. Arthaud, F. Deliot, B. Tuchming, V. Sharyy and D. Vilanova, Muon momemtum oversmearing for p17 data, DØ Note 5444 (2007).

[63] P. Calfayan *et al.*, Muon identification certification for p17 data, DØ Note 5157 (2006).

[64] P. Calfayan, Oring single muon triggers in p17 data, DØ Note 5329 (2008).

[65] DØ Collaboration, Search for second generation scalar leptoquarks in the $\mu\nu jj$ final state with the DØ detector, Conference Note 5370 (2007).

[66] A. Höcker *et al.*, TMVA: Toolkit for multivariate data analysis, (2007), physics/0703039 (version v3.8.14 has been used for the search presented in this thesis).

[67] DØ Collaboration, T. Andeen *et al.*, FERMILAB-TM-2365.

[68] D. Cho *et al.*, Measurement of the $t\bar{t}$ production cross section at $\sqrt{s} = 1.96\,\text{tev}$ in the lepton+jets final state using kinematic information on $1\,\text{fb}^{-1}$ of data, DØ Note 5527 (2008).

[69] R. C. P. Bartalini and A. D. Roeck, CMS-NOTE-2005-013 **A434** (2005).

[70] V. Büscher, J.-F. Grivaz, T. Nunnemann and M. Wobish, Conclusion of mini-workshop on pdf uncertainties and related topics, DØ Note 4618 (2004).

[71] T. Junk, Nucl. Instrum. Meth. **A434**, 435 (1999), [hep-ex/9902006].

[72] R. Brun and F. Rademakers, An object oriented data analysis framework (http://root.cern.ch).

[73] J. Neyman and E. S. Pearson, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231**, 289 (1933).

[74] http://www-d0.fnal.gov/∼nunne/cross-sections/lq_cross-sections.html.

[75] DØ Collaboration, B. Abbott *et al.*, Phys. Rev. Lett. **83**, 2896 (1999), [hep-ex/9904023].